Cerebral Function and Connectivity in Twins with Bipolar Disorder.

Fergus Kane, BSc MSc

Institute of Psychiatry, King's College London

University of London

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy in the University of London

September 2008

Dedication.

This thesis is dedicated to Katherine Sheldon.

Acknowledgements.

Firstly, I would like to thank the many volunteers who selflessly gave the gift of their time and effort in order to further our understanding of Bipolar Disorder. It has been a true pleasure to work with so many interesting and dedicated people.

I would like to acknowledge my supervisors, Professor Robin Murray and Professor Philip McGuire for their help and advice and for making this thesis possible.

I wish to acknowledge all those who, despite having no official role in my supervision, took the time to provide invaluable help and support. I wish especially to thank Professor Gareth Barker, Dr Andrea Mechelli, Dr Simone Reinders, Dr Vincent Giampietro and Dr Christopher Chaddock. Thanks go also to the many other people who have made small but significant contributions along the way.

Special thanks goes to my colleagues on the Maudsley Bipolar Twin Study: Dr Eugenia (Jenny) Kravariti, Dr Sridevi Kalidindi and Anna Georgiades. To Jenny: without your intervention, this thesis would not exist and for that, I am deeply grateful.

Finally, to my family. Without you, I would never have had the opportunity to begin, let alone complete this work. Your support, both tacit and unyielding, lies too often unacknowledged. You have provided me with the space to find my own path, a precious freedom, available to so few. My gratitude, though rarely stated, is profound. Thank you.

Work Completed By Candidate.

Work on the Maudsley Bipolar Twin Study, of which this thesis forms a part, was carried out by four researchers: Fergus Kane, Dr Eugenia Kravariti, Dr Sridevi Kalidindi and Ms Anna Georgiades. Specific responsibilities are detailed below:

Recruitment and Testing.

Recruitment for the study was conducted jointly between Fergus Kane, Dr Kalidindi, Dr Kravariti and Ms Georgiades.

Clinical Assessments and DNA:

Clinical Assessments were conducted by Fergus Kane, Dr Kalidindi and Ms Anna Georgiades. The above researchers were also responsible for the collection of blood sample and cheek swabs for DNA analysis and zygosity testing.

Neuropsychological Testing

Neuropsychological testing for the study was conducted jointly by Fergus Kane, Dr Kravariti and Ms Georgiades.

MRI Imaging

MRI imaging for the study was conducted by Fergus Kane and Dr Kalidindi in conjunction with Institute of Psychiatry Radiographers.

Data Management

Fergus Kane was responsible for data management for the study.

Analysis

The candidate conducted all the analysis found in this thesis. Professor Gareth Barker provided guidance for DTI analysis. Dr Andrea Mechelli provided guidance for fMRI and connectivity analysis. Dr Simone Reinders provided guidance on MATLAB programming. Dr Vincent Giampietro provided advice on the XBAM brain imaging package. Daniel Stahl provided advice on statistical analysis in STATA.

Abstract

Despite considerable recent research, the aetiology of bipolar disorder remains largely unexplained. From twin studies, it is known that bipolar disorder has a strong genetic component. However, the search for genes involved in bipolar disorder has been less fruitful than was originally expected. It is now widely accepted that this is due to the fact that bipolar disorder is likely to be influenced by many genes of small effect, rather than a few genes of large effect.

It has been proposed that the characterisation of endophenotypes (alternative, well defined elementary phenotypes that are more closely linked to the genotype than the clinical phenotype) may provide further insight into the genetic basis of the disorder. This thesis describes my body of work, which sought to investigate previously reported abnormalities of neural function and white matter, and to assess their potential as endophenotypes for bipolar disorder, using both functional Magnetic Resonance Imaging and Diffusion Tensor Imaging. I obtained neuroimaging data for 112 subjects, comprising identical and fraternal twin pairs both concordant and discordant for bipolar disorder as well as control twin pairs. Where abnormalities were confirmed within this sample, the study also explored the extent to which they lay under genetic control – and were thus suitable as candidate endophenotypes.

Patients with bipolar disorder did not demonstrate any abnormalities of neural activity as measured via fMRI. However, patients did demonstrate abnormalities of white matter tracts within the genu, body and splenium of the corpus callosum, as well as in the inferior and superior longitudinal fasciculus. Further, these abnormalities were also seen in the unaffected co-twins of patients with bipolar disorder. My results indicate that abnormalities of white matter represent a potential endophenotype for bipolar disorder. The identification of white matter deficits as a potential endophenotype will assist further understanding of the aetiology of bipolar disorder. Pending further confirmation, the identified white matter abnormalities may be used as quantitative phenotypes for the identification of susceptibility genes for bipolar disorder.

Table of Contents

List of Tables

List of Figures

1. Introduction

This chapter aims to: (i) Introduce the reader to the concept of bipolar disorder, (ii) introduce the Maudsley Bipolar Twin Study, (iii) explain the aims of the present thesis, and (iv) provide a representative overview of the relevant literature. Please note, while all abbreviations are introduced in the text, the reader is invited to refer to list of abbreviations found in appendix B, in case of any confusion.

1.1. An Introduction to Bipolar Disorder

Bipolar disorder (BD) is a highly disabling dysfunction of mental state, which is characterised by unusual intensity and lability of mood. Based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM IV)[1], the disorder is commonly divided into two subtypes: bipolar disorder type I (BD-I) and bipolar disorder type II (BD-II). Both are characterised primarily by the presence of manic or hypomanic episodes, which are defined as 'distinct periods of abnormally and persistently elevated, expansive, or irritable mood'. Typical symptoms of a manic episode include inflated self-esteem or grandiosity, decreased need for sleep, unusual talkativeness or pressure of thought, flight of ideas, distractibility, increases in goal directed activity and excessive involvement in pleasurable activities. Hypomanic episodes are similar to manic episodes but are of lesser severity, generally with less adverse consequences. To merit a diagnosis of BD-I, an individual must have experienced one or more manic or mixed episodes, while to reach a diagnosis of BD-II, at least one hypomanic episode AND at least one major depressive episode is required. More broadly the 'bipolar spectrum' also includes 'cyclothymia' which is defined by a period of at least two years, during which an individual has experienced numerous periods of subthreshold hypomanic and depressive symptoms (which do not meet the criteria for either major depressive or manic episodes). For a comprehensive definition of bipolar disorder and its subtypes, the reader is referred to the Diagnostic and Statistical Manual of Mental Disorders[1].

The lifetime prevalence of BD-I in the general population has been estimated at 1% and is relatively consistent across nations, race and socioeconomic status [2]. However, more recent estimates that include the full spectrum of bipolar disorders report a prevalence of 3-5%. Age of onset is typically early 20s in males, slightly later in females, but adolescent and late onsets are also seen[3,4]. Heritability estimates for bipolar disorder, as ascertained from family and twin studies vary from 50% to above 80% [5].

Bipolar disorder, almost invariably, causes severe disruption to the lives of those it affects. Inevitably, this disruption extends to the lives of a sufferer's relatives, friends and colleagues. The chronic, yet episodic, nature of the disorder means that maintenance of careers and relationships is particularly difficult. Approximately 10-20% of individuals with bipolar disorder commit suicide, while approximately one third of patients admit to at least one suicide attempt [6]. The social and economic costs of bipolar disorder are also considerable; a 1991 estimate put the economic cost for the US alone at $45 billion per year [7]. Indeed, the severity of bipolar disorder is reflected in the World Health Organisation's (WHO) oft cited estimation that BD is (in terms of years of life lost due to a disability) the sixth most disabling illness worldwide[8]. Given these facts, it is clear that any improvement in our ability to treat this disruptive disorder would have significant benefits, both for individuals with bipolar disorder and their families, as well as for society as a whole. Our knowledge of the aetiology of this disorder remains limited and clearly more research is needed in order that we may more fully understand it.

Bipolar disorder has been a recognised as a psychiatric illness since at least the late 19th century, when Emil Kraeplin first distinguished between manic-depressive insanity (bipolar disorder) and dementia praecox (schizophrenia). Despite this, traditionally, research into bipolar disorder has been scarce compared to research into schizophrenia. However, increasing recognition of the considerable burden of this disease, both to individuals and to society, means that research aimed at elucidating its aetiology now has a higher priority than before. This thesis, and the larger Maudsley Bipolar Twin Study of which it is part, form a part of, and benefit from, this new wave of research into bipolar disorder.

1.2. The Maudsley Bipolar Twin Study and the Present Thesis:

1.2.1. The Maudsley Bipolar Twin Study

The Maudsley Bipolar Twin Study (MBTS) is part of a tradition of twin research at the Institute of Psychiatry, London. In particular it builds upon the work carried out as part of the 'Maudsley Twin Register', which was initiated in 1948 by Slater. The MBTS itself was initiated in 2003, in order to investigate potential endophenotypes of bipolar disorder. To this end, the study has recruited (and continues to recruit) twins with a diagnosis of BD from around the UK. Participants in the study participate in a large range of neuropsychological and neuroimaging tasks, as well as providing genetic data for both zygosity testing and genetic analysis.

1.2.2. This Thesis

The current thesis presents the findings from the neuroimaging section of the MBTS. Specifically, data is presented from:

1. A functional magnetic resonance imaging (fMRI) study of working memory in BD
2. A diffusion tensor imaging (DTI) study of white matter in BD.

The long term aim of the MBTS is to use twin modelling techniques (involving structural equation modelling) in order to investigate potential endophenotypes. However, despite four years of recruitment, due to the rarity of the sample, it has not been possible to carry out brain scans on a sample of sufficient size to reliably perform a structural equation modelling analysis. Therefore, the current thesis presents data from more traditional analyses, without attempting to calculate heritability estimates. Recruitment and study design are described in more detail in the methods section of this thesis.

1.2.3. The Endophenotype Concept.

Standardised psychiatric diagnoses have evolved primarily as a way of categorising people according to the symptom clusters that they present with. This has led to diagnoses that describe the most commonly observed presentations. However, in reality, all patients present with slightly different symptoms, thus the definitions of each disorder have been designed to accommodate a wide range of symptoms – some of which may be present and some of which may be absent. The resulting heterogeneity of presentations is problematic for psychiatric researchers, in that it may reduce the power of a study to characterise the biological underpinnings of disease. Furthermore, not only are the standard psychiatric phenotypes too broad for focussed research, there is also increasing evidence that diseases such as schizophrenia, bipolar disorder and depression have significant shared genetic vunerabillity[9-11].

It has been suggested therefore, that in order to increase the power of studies aiming to reveal the biological and genetic underpinnings of psychiatric disorder, it is necessary to explore alternative, more circumscribed phenotypes. There are two main approaches that have been adopted to this end; the first involves a refinement of the standard phenotypes by using a symptom dimensions or symptom cluster approach (for an example, see Boks et al.[12]), the second is the use of the endophenotype concept, and it is this approach that is discussed hereafter.

It has been suggested that individual susceptibility genes related to illnesses such as bipolar disorder and schizophrenia produce distinct, circumscribed phenotypes that may each increase the risk for the disorder [13]. These well-defined elementary phenotypes are often called endophenotypes, intermediate phenotypes, trait markers or vulnerability markers. The endophenotype may be more closely linked to the genotype than the disease phenotype and thus identification of endophenotypes may provide a powerful tool for investigating the biological basis the disease. In an influential 2003 paper, Gottesman and Gould defined five criteria for endophenotypes[14]:

1. The endophenotype is associated with the illness.
2. The endophenotype is heritable.
3. The endophenotype can be detected in remitted patients who do not suffer from active illness (state-independence).

4. The endophenotype and the illness co-segregate in affected families.

5. The endophenotype found in ill family members is found in unaffected family members at a higher rate than in the general population.

While the concept of the endophenotype has been available to psychiatry for over thirty years, it is only since the 1990s that psychiatric endophenotypes have been widely investigated[15]. Much of the work conducted to date has been in the field of schizophrenia, and in this disorder, some of the putative endophenotypes do appear to match these criteria. For instance, cognitive deficits in schizophrenia emerge significantly earlier than clinical symptoms[16,17], are present both during episodic exacerbations and clinical remission [18]; have been proposed to be sufficiently reliable to form diagnostic criteria [19]; and, importantly, are also seen in an attenuated form in unaffected relatives of schizophrenia patients [20-22], including children at high familial risk for developing the disorder [23].

Endophenotypes may take a variety of forms, as long as they fulfil Gottesman and Gould's five basic criteria. In bipolar disorder, observed abnormalities of cognition, event related potentials (ERPs), brain structure, brain connectivity and brain activity (as measured by functional magnetic resonance imaging, fMRI) have all been suggested as potential endophenotypes. However, it is currently unclear which, if any of the abnormalities reported in bipolar disorder may serve as markers of genetic vulnerability to the disorder.

Establishing such markers can help: (a) identify individuals at increased genetic risk, with important implications for prevention and early intervention; (b) identify susceptibility genes; (c) improve our understanding of the biological underpinnings of the disorder. The current state of this evidence is discussed in more detail later. Firstly, however, it important to note that the idea of using endophenotypes as a tool for identification of susceptibility genes is not without its difficulties, these are discussed below.

Criticism of the Endophenotype Concept as a Tool in Psychiatric Genetics

The basic logic behind using endophenotypes to identify susceptibility genes is that, following successful identification of an endophenotype for a disorder, follow-up studies would be run using the endophenotype (in place of the standard disease phenotype) to leverage the search for candidate genes. There are a number of reasons why using endophenotypes might be advantageous relative to the standard disease phenotype. For instance, endophenotypes may be easier and cheaper to measure than the phenotype as well

as being potentially more reliable. However, the concept of using endophenotypes to identify susceptibility genes is predicated on the idea that endophenotypes are more 'defined and quantifiable measures' than disease phenotypes, with a simpler path from the genotype to endophenotype than the disease phenotype[14]. If the later is true, and there is a simpler path, variations in the DNA sequence should interact more directly with the endophenotype, resulting in greater correlations that are thus easier to detect than those between the standard phenotype and the genetic variations. This idea, according to Flint and Munafo[15], is an assumption that has been largely untested in psychiatric disorders. Indeed, although the endophenotype approach has helped to identify susceptibility genes in non-psychiatric disorders such as the long-QT cardiac syndrome[24], in the psychiatric arena, it has yet to live up to early expectations.

In order to investigate the assumption that endophenotypes would provide larger effect sizes than disease phenotypes, Flint and Munafo conducted a meta-analyses of gene-endophenotype association studies in schizophrenia and compared these to their own gene-disease phenotype meta-analysis. This was done for one gene (COMT) and three suggested endophenotypes, WCST performance, N-Back performance and p300 amplitude. In all three cases, the effect sizes (1.1, 1.1 and 1.01 respectively) reported for the putative endophenotypes were less than for the (non-significant) schizophrenia phenotype itself (1.13). If this were to hold true for other genes and endophenotypes, the sample sizes needed to detect genetic loci for endophenotypes, would be no smaller than those for disease phenotypes. Given that a major justification for the search for endophenotypes is that they may confer greater statistical power, this is a significant issue. Furthermore, these results also challenge the assumption that endophenotypes have a simpler genetic architecture than their associated disease phenotypes (as this should result in greater effect sizes). As the authors admit, this study is limited by the fact that only one genetic variation was investigated. Equally, it is possible that the endophenotypes chosen had a particularly complex genetic architecture and that other genes and other endophenotypes may demonstrate stronger associations. However, the author's primary point remains, it is far from certain whether, in disorders such as bipolar disorder, endophenotypes can provide a significantly more powerful way of identifying susceptibility genes than disease phenotypes themselves.

Flint and Mufano's criticism is an important reminder that the search for endophenotypes is unlikely to serve as a panacea in the quest to improve our understanding of psychiatric disorders. Despite this, it important not to loose sight of the potential benefits of the endophenotype approach. As noted earlier, endophenotypes are potentially much easier to measure and may be more reliable than standard psychiatric phenotypes and thus may facilitate relatively quick recruitment and testing of the large samples required for genetic studies. Thus, even if the effect sizes for endophenotypes are not significantly higher than for the phenotype, they may still have utility in the search for susceptibility genes. Equally, the identification of endophenotypes may help to refine existing animal models and/or identify novel targets for psychological and pharmacological treatments. Finally, the identification of endophenotypes may be considered an important goal in itself, as true endophenotypes should provide insights into the aetiology and mechanisms underlying bipolar disorder.

1.2.4. The Utility of Twin Methodology in the Search for Candidate Endophenotypes

Twin studies are a particularly powerful way of disentangling genetic, environmental and disease related factors governing brain function and behaviour. In particular, studies of monozygotic (MZ) and dizygotic (DZ) twin pairs discordant for bipolar disorder enhance the sensitivity to detect neurocognitive abnormalities, because shared genetic (MZ=100%, DZ=~50%), prenatal and postnatal environmental factors that contribute to inter-individual variation are controlled for. In such research, a genetic role is suggested when MZ patients and their co-twins differ from healthy MZ twins but do not differ from each other, and the finding is more pronounced in discordant MZ than discordant DZ pairs.

Twin (and family) studies have the additional advantage that they allow researchers to address the confounds of medication and sub-syndromal symptoms; this is discussed in more detail later in the chapter.

1.3. Cognitive Dysfunction in Bipolar – Endophenotypes?

Deficits in cognitive function have been associated with bipolar disorder and these deficits have been reported during mania, depression and in symptom remission. The fact that such deficits are both associated with bipolar disorder and are state independent, suggests that they may represent potential endophenotypes (criteria 1 and 3 from Gottesman and Gould).

To date, more than 70 studies have investigated cognitive function in euthymic disorder. The majority have used small sample sizes (often including fewer than 30 patients), a caveat which restricts both their power to detect effects and their generalisability to the wider patient population. Three recent meta-analyses[25-27] have attempted to overcome this problem.

The first of these meta-analyses, by Robinson et al[26], pooled and quantified neurocognitive evidence from 26 independent studies of euthymic bipolar patients and controls published between 1980 and August 2005. Sixteen cognitive variables (each used in at least four studies), as well as years of education, were included in the meta-analysis. Overall, bipolar probands performed significantly worse than controls on all variables except IQ (and years of education). Effect sizes (categorised according to Cohen's convention[28]) varied from small (0.2-0.49) to medium (0.5-0.79) to large (0.8+). In the executive domain, category fluency and digit span backward (mental manipulation/working memory) showed large effect sizes, stroop performance (response inhibition and susceptibility to interference) and Wisconsin Card Sorting Test (set-shifting and abstraction) gave rise to medium effects, while verbal fluency had a small effect. In the memory domain, a large effect was seen for immediate verbal recall/learning, medium effects for short- and long-delay free verbal recall, and a small effect for digit span forward. Finally, in the attention/psychomotor speed domain, medium effect sizes were seen for 3 variables (response latency in sustained attention tasks, digit symbol substitution and trail making A), and a small effect for a fourth variable - sustained attention sensitivity. Robinson et al[26] found no evidence of publication bias in the available literature. Two later meta-analyses by Arts et al[27] and Torres et al[25] largely confirmed the findings by Robinson et al[26].

Table 1.1 summarises the effect sizes estimated by the three meta-analyses for various components of intelligence, verbal memory/learning, attention/psychomotor speed, and executive function (Arts et al[27] further estimated effect sizes for first-degree relatives of bipolar patients – also included in Table 1). It should be noted that, as the three systematic reviews inevitably included many of the same studies, they should be considered as confirmatory analyses rather than as replication studies. Taken together, the three meta-analyses provide compelling evidence that deficits of cognitive function do indeed exist in euthymic disorder. However, it remains possible that such deficits are due to confounds such as sub-syndomal mood symptoms and medication[1].

Even if a cognitive deficit is present in euthymic patients, this is not enough to confer upon it endophenotypic status. It is still necessary to show that the deficit is heritable, that the deficit and the illness co-segregate in affected families and that the deficit found in ill family members is also found in unaffected family members at a higher rate than in the general population. These questions can only be answered using family and twin studies. Such studies also provide the advantage of being able to exclude the confounds of medication and (to some extent), sub-syndromal symptoms (discussed later in this chapter).

The family and twin literature in bipolar disorder presents a very mixed picture, characterised more by negative than positive findings. While many of the studies have reported specific cognitive deficits in the unaffected relatives of bipolar patients, these have not been consistent between studies and for each variable investigated there have been more non-significant than significant findings. Again, meta-analytic techniques have been employed in order to make sense of these disparate findings. Arts et al[27] found that, in the non-bipolar relatives of bipolar patients there was evidence for cognitive deficits in five out of the twelve variables that they had included in their meta-analysis. Four of these variables (stroop performance, verbal fluency, trails-B and immediate verbal recall) had small effect sizes, while the remaining variable (long delay verbal recall) had a medium effect size. The remaining variables (Wisconsin Card Sorting Test categories and perseverative errors, digit

[1] The potential neuropsychological effects of sub-syndromal symptoms and mediation are discussed later in this chapter.

span forward, trails A, digit symbol substitution, digit span backward, IQ and letter fluency) had effect sizes of less than 0.2. Furthermore, out of all of these variables, only immediate verbal recall, stroop and trails-B demonstrated statistically significant differences in the meta analysis.

Beyond the meta-analyses detailed above, there are a small number of narrative and systematic reviews of the evidence for endophenotypes in bipolar disorder. Balanza-Martinez et al[29] recently published a systematic review that looked at 23 studies of cognitive function in the relatives of patients with bipolar disorder. Of the studies investigated, 6 of 11 reported deficits in verbal learning and memory, 3 or 9 in working memory, 1 of 6 in visuo spatial learning, 1 of 8 in alternating attention, 2 of 10 in psychomotor speed, 2 of 8 in cognitive flexibility or attention, 2 of 16 in general intelligence and 0 of 6 reported deficits in immediate memory or verbal fluency. The authors conclude that the neurocognitive profile of bipolar disorder remains unclear, but that the best evidence of cognitive deficits lies in the domains of verbal learning and memory, which are perhaps most pronounced in long delayed recall, but are also seen in working memory. Deficits of long delayed recall also seem to be evident on list learning, but not story learning tasks, indicating a specific deficit in executive strategy underlying list learning. Deficits in working memory were observed in one third of studies, but more specific deficits of verbal working memory were observed in half (3 of 6) of these studies that examined this. The authors also note as remarkable the fact that IQ appears to be spared in relatives of patients, although, given that there is very little evidence for IQ deficits in patients, this is not altogether surprising. These findings and conclusions roughly reflect those of the meta-analysis of Arts et al[27], although they differ in that Arts et al found deficits in verbal immediate recall and trail making, both of which were not prominent in the systematic review. The systematic review also indicated potential differences in verbal working memory (on WAIS digit span forward), which were not significant in the meta-analysis. Here, Balanza-Martinez et al report an important flaw in the meta-analysis of Arts et al. It appears that in the above cases, Arts et al have double counted subjects from non-independent samples, possibly resulting in errors in weighted effect sizes. This is, unfortunately, a not uncommon error in meta-analyses[30] and highlights the importance of referring to the individual studies. Nevertheless, the studies complement each other and their conclusions are similar, that there remains limited evidence from the relative literature for endophenotypes, the strongest of which appears to be in the domains of verbal memory, learning and working memory. Finally Balanza-Martinez et al make the important point that

the literature is lacking in studies investigating other potentially important functions such as language, social cognition and planning and motor skills.

Another recent qualitative (but not systematic), review[31] of the evidence for cognitive endophenotypes in bipolar disorder largely supports the findings of the above study. This study looked at the meta-analytic evidence from studies of cognition in patients as well as reviewing key papers on cognitive deficits in relatives of patients. The study found that while there was reasonable evidence (based on the published meta-analyses) of cognitive dysfunction in bipolar disorder, the evidence that these represent endophenotypes for the disorder was less clear, being strongest within the domains of verbal memory and executive function. In particular, the study found that for most neuropsychological variables investigated in the relatives of patients with bipolar disorder, there were considerably more negative that positive findings. This however, cannot be taken to mean that these variables can be ruled out as potential endophenotypes. As the authors point out, Clark et al[32] suggest that in order to confirm group differences with a small/medium effect size of 0.38, studies should have at least 115 participants in each group. All studies to date fall short of this aim, most by a considerable margin. While the problem of small samples can be overcome to some extent by meta-analytic techniques such as those already discussed, there remains a need for larger individual studies.

These conclusions of both the above reviews are also broadly similar to those from an earlier review by Glahn et al[33]. This review looked at the published evidence available in late 2003 and highlighted verbal learning, verbal memory, executive function and working memory as the most convincing cognitive targets for endophenotype research. Thus, over the last half-decade, the strongest candidates for cognitive endophenotypes have remained similar.

While Balanza-Martinez et al have provided a systematic review of evidence for cognitive dysfunction in the relatives of patients with bipolar disorder; there is perhaps surprisingly, no such systematic review of the evidence in patients themselves. In 2006, Robinson and Ferrier[34] published a systematic review, which rather than concentrate on the evidence for cognitive deficits in bipolar disorder, looked at the clinical correlates of cognitive impairment in patients with bipolar disorder. The key finding of the paper was 'a robust association between impaired long-delay verbal memory and a greater burden of illness'. This finding was particularly strong for number of manic episodes, which was negatively correlated with long delay verbal memory in four of six papers that looked for such a relationship. The

authors speculate that an underlying executive function deficit may be responsible for verbal memory dysfunction, but that this may initially be compensated for by way of alterative cognitive strategies. While it is thus unclear which is the primary dysfunction, this finding adds weight to the idea that deficits of verbal memory and executive functions are important in bipolar disorder.

Overall then, there is limited evidence in the bipolar literature to support cognitive deficits as endophenotypes for bipolar disorder; of these, specific impairments of verbal memory and executive function present the strongest candidates to date. These domains are known to be highly heritable[35-37], they appear to be associated with bipolar disorder, appear to be independent of clinical state[38,39] and evidence suggests deficits in these domains are present in unaffected relatives. As a result, these cognitive functions present perhaps the most obvious candidates for investigation using neuroimaging techniques.

Table 1.1 Meta-analyses of Cognitive Function in Euthymic Bipolar Patients and Non-Bipolar Relatives

	Effect Sizes (Cohen's d (# studies))			
	Euthmic Patients			Relatives
	Torres et al	Robinson et al	Arts et al	Arts et al
IQ Measures				
IQ		0.19 (12)	0.16 (8)	0.19 (5)
Reading	0.04 (19)			
Vocabulary	0.08 (10)			
Verbal Learning / Memory				
Immediate Recall/Learning	0.81 (12)	0.90 (10)	0.82 (12)	0.42 (4)
Short Delay	0.74 (10)	0.73 (10)		
Long Delay	0.72 (12)	0.71 (11)		
Delay (unspecified)			0.85 (10)	0.56 (4)
Recognition hits	0.43 (10)			
Digit Span Forward		0.47 (5)	0.37 (6)	0.04 (4)
Attention/Psychomotor Speed				
Trails A	0.60 (10)	0.52 (11)	0.71 (10)	0.13 (7)
Sustained Attention				
Hits	0.74 (8)			
Sensitivity		0.48 (4)	0.58 (4)	
Reaction Time	0.62 (10)	0.60 (7)		
Digit Symbol Substitution Test	0.79 (8)	0.59 (9)	0.84 (7)	0.14 (4)
Executive Function				
Digit Span Backward	0.54 (8)	0.98 (5)	1.02 (6)	0.18 (5)
Trails B	0.55 (11)	0.78 (12)	0.99 (10)	0.37 (7)
Wisconsin Card Sorting Test				
Categories	0.69 (8)	0.62 (7)	0.52 (10)	0.04 (4)
Perseverative Errors		0.76 (7)	0.88 (10)	0.17 (6)
Fluency (categories)		1.09 (4)	0.87 (7)	
Fluency (FAS)	0.47 (11)	0.34 (8)	0.59 (12)	0.27 (4)
Stroop				
All Measures	0.71 (13)	0.63 (11)		0.49 (4)
Correct			0.65 (8)	
Time			0.73 (6)	
Rey Figure Recall			0.62 (4)	
Rey Copy			0.22 (4)	

Effect size coding: red=large, green=medium, blue=small

1.4. Neuroimaging in Bipolar Disorder

The following section details the neuroimaging research most relevant to bipolar disorder. In order that the discussion of this research is clear, concepts that are common to all neuroimaging modalities are briefly introduced below. Concepts that are specific to a particular modality of neuroimaging will be introduced in relevant sections.

1.4.1. A Very Brief Primer on Neuroimaging and Neuroimaging Concepts.

Neuroimaging is the imaging of the brain using (generally) non-invasive techniques. The earliest form of neuroimaging used X-ray tomography, a technique widely used since the late 1970s. Since then a number of other techniques have been developed, including PET (positron emission tomography), SPET (single photon emission tomography) and MRI (magnetic resonance imaging). Neuroimaging is very much in its infancy, but has already proved to be invaluable both from a clinical and a research perspective. Neuroimaging research into bipolar disorder has mirrored neuropsychological research in that there has been a relative paucity of publications when compared to schizophrenia, but the upswing in bipolar disorder research means that there is now a sizeable literature, in PET, SPET and particularly in MRI.

Data from neuroimaging studies can be analysed in a number of ways, but the most common of these are region of interest (ROI) analyses and whole brain analyses. The ROI method involves using a-priori hypotheses to select specific regions of the brain that are believed to differ between diagnostic groups. Typically, for each subject, a trained neuroanatomist will select these regions on the brain images of each subject and the data from these regions will be extracted for further analysis[ii]. Whole brain analysis methods, by contrast, do not normally use a-priori hypotheses. Instead, group brain maps are created from individual scans and these group maps are compared voxel by voxel[iii].

Both of the above methods have attendant advantages and disadvantages. ROI methods offer potentially greater statistical power than whole brain methods, but this comes at the risk of missing important differences in regions that are not included in the a-priori regions. ROI

[ii] For reliability reasons, these regions will often also be compared to the same regions selected by a second neuroanatomist.

[iii] A voxel (a portmanteau of volume pixel) is a data point in a 3D image.

studies are also very labour intensive; the process of accurately selecting three-dimensional ROIs is time consuming and difficult. The reliable delineation of regions of interest is also an issue in ROI studies. While for some clearly defined areas of the brain (eg. the corpus callosum, cerebellum), reliable region delineation is relatively straightforward, for other areas it is more difficult and thus also more likely to be subject to systematic biases. Reliable delineation may be challenging for various reasons, but may be especially difficult where contrast between different tissues is limited, where regions are small and are at the limits of the image resolution, where there is considerable intersubject variability or where region boundaries are not clearly defined theoretically. Furthermore, it is quite possible that while there may be no detectable differences in the size or activation of a particular region, there may be important differences in the region's subunits. While such subunits may be measured as separate ROIs, due to reduced tissue contrast and smaller size, accurate delimitation of subunits will often be more difficult than for the whole structure.

While brain analyses, by their nature, investigate every voxel of the brain and thus can overcome many of the above problems. However, due to the number of comparisons required to compare the group maps at every voxel in the brain, risk type I errors. Such errors can be accounted for by adjustment for multiple comparison correction, but this means increasing the risk of type II errors. Furthermore, whole brain analyses involve a number of extra steps (such as image registration, normalisation and smoothing) that can significantly affect the results of the analysis. Whole brain analyses are thus most suitable for exploratory studies, where strong a-priori hypotheses are not available. Conversely, ROI studies are most suited to studies where previous research points towards a very specific hypothesis

1.4.2. Structural Imaging in Bipolar Disorder.

White Matter Hyperintensities

It is not clear what causes the neuropsychological and clinical abnormalities seen in bipolar disorder. Evidence from structural neuroimaging, while not entirely consistent, suggests that these abnormities may be related to white patter pathology. Early evidence comes from the observation of elevated levels of white matter hyperintensities in bipolar disorder, an observation that has been found fairly constantly over the last two decades[40-45]. White matter hyperintensities are areas of increased T_2-weighted MRI signal best that are associated with a variety of pathology including axonal loss, astrogliosis and demyelination[46]. At a functional level such hyperintensities have been associated with suicidality[47] and poor outcome[48]. These associations have led to implication of white matter disease in the aetiology of bipolar disorder. Hyperintensities are observed in the normal population, but are only generally seen in later life, by contrast, in bipolar patients they can be observed as early as adolescence[49]. Nevertheless the presence of hyperintensities in the normal population, as well as the lack of hyperintensities in a significant proportion of bipolar patients[50,51] means that their role in the disorder remains unclear.

Volumetric Changes

Region of Interest Studies

Studies investigating anatomical abnormalities in bipolar disorder have yielded mixed results. The most consistent findings appear to be of unaffected whole brain volume[52,53] and increased ventricle-brain ratios[52,54-57]. Findings from studies investigating smaller or less easily defined areas are more complicated. According to a recent review by Konarski et al[58], volumetric abnormalities of the temporal lobe, hippocampus, amygdala, basal ganglia, cerebellum, thalamus and pituitary have all been reported. However, for many of these regions, reports are conflicting; for instance almost as many studies have reported enlargement of the amygdala as have reported reductions.

Most recently Kempton et al[59] have published a meta-analysis of 141 studies investigating brain structure in bipolar disorder. The meta-analysis investigated brain regions where results had been reported in at least 3 papers, which yielded 47 regions. Of these 47 regions, only the lateral ventricles (total and right side, but not left), third

ventricle and corpus callosum (cross sectional area) showed any bipolar-control differences; ventricular size was enlarged by between 9.6% and 24.2% (es = .16 to .52), while the corpus callosum was reduced by 7.3% (es =-.43). There was no evidence of publication bias for these areas. The study also investigated white matter hyperintensities and found evidence of increased deep white matter and sub cortical grey matter hyperintensities as well as increased hyperintensities in the both hemispheres, frontal and parietal lobes. However, with regard to hyperintensities, evidence of publication bias was found for combined and right hemisphere variables. The findings of this meta-analysis are largely what one might expect given the lack of aforementioned lack of consistency in the literature. As the authors note, it remains unclear whether the findings reflect a genuine lack of structural change in bipolar disorder, or whether they can be ascribed to between study heterogeneity. It is probable that the uncertainty in the field may only be overcome by very large and well-controlled studies.

Voxel Based Morphometry Studies

In addition to the ROI studies briefly discussed above and summarised in the meta-analysis of Kempton et al[59], there is also a growing literature of whole brain, Voxel Based Morphometry (VBM) studies in bipolar disorder and this is discussed below.

It may help to briefly introduce some common terminology used in VBM. VBM studies generally report either differences in 'density' or in 'volume' (or both). These variables arise from the process of spatially normalising individual subject's data to a standard template (whereby analysis software attempts to warp individual subject's brain images to match a standard template, so that they may be more reliably compared).

In VBM, density and volume are related but not equivalent. A voxel's density, which is represented as a value between 0 and 1, can be thought of as the probability that the voxel represents a particular brain tissue (e.g. grey matter), or as the proportion of the voxel that is made up of a particular brain tissue. A voxel's density value however, does not make use off all the information we have about that voxel. This is because nonlinear spatial normalisation of brain images to a template inevitably results in different brain regions experiencing differential volume changes and information about such changes is not encoded in the density maps. In order to integrate this

information, the density data is modulated using the parameters from the normalisation step; that is to say, it is scaled so that the total amount of grey matter remains the same as in the original images. When VBM studies refer to 'volume' differences between groups, they are referring to differences in this modulated data. It is important to note that VBM does not provide information about the absolute volume of different brain areas.

The majority of the VBM studies in bipolar disorder have been case-control studies and the results of these studies are summarised below. Two of the VBM studies in bipolar disorder have instead chosen to look at the relationship between brain morphology and genetic liability to bipolar disorder. As these are of particular interest due to their relevance to the endophenotype concept, they are discussed in more detail. The results of all the studies are summarised in Table 1.2.

In the case-control studies, increased grey matter volume has been reported in the left insula/frontoparietal operculum[60], ventral occipitotemporal cortex[60], basal ganglia[61], anterior cingulate[62], ventral prefrontal cortex[62], fusiform gyrus[62,63], motor cortex[62], thalamus[63], cerebellum[63], middle and superior temporal cortex[63], paracentral lobule[63], parahippocampal cortex[64,65] and hippocampus[65].

Decreased grey matter volume has been reported in temporal cortex (inferior[66,67], middle[67,64], superior[66] and medial[60,61]), orbito frontal cortex[61], cingulate (posterior[66,60,67] and anterior[60,61]), superior parietal lobe[62], frontal cortex (inferior[67], ventral prefrontal[66], posteromedial gyrus rectus[68]), insula[67], parahippocampal gyrus[68], precentral gyrus[67] and left putamen[68].

Grey matter density has been reported in four studies. Increased density has been found in cingulate (anterior[69,63] and posterior[63]), mid temporal[63], parietal lobe (inferior[63] and superior[63]), precuneus[63], precental gyrus[69,63], fusiform gyrus[63], frontal cortex (medial[69] and inferior[69]), frontoparietal cortex[60], insula[60] and thalamic cortex[60]. Decreased grey matter density has been reported in ventromedial temporal cortex[60], anterior thalamus[70] and caudate[70].

Only four of the published case-control VBM studies have investigated white matter differences, three in patients and one in patients and unaffected relatives. Of these, one reported decreased white matter density in bilateral prefrontal cortex[71], one reported decreased white mater volume in left medial parietal cortex (in unmedicated BD patients) and increased left ventral prefrontal cortex volume (in medicated BD patients)[66] and one reported decreased white matter volume in left frontal cortex and bilateral temporoparietal junction[67]. The fourth paper[72] investigated patients and unaffected relatives from both bipolar families and mixed bipolar and schizophrenic families (separately). This paper reported that patients from bipolar families, but not those from mixed families showed reductions of white matter in the anterior limb of the internal capsule. Unaffected relatives from mixed, but not from bipolar families showed reductions of white matter in the right superior frontal and medial frontal gyri.

As noted earlier, two of the published VBM studies are of particular interest, because they have attempted to investigate the relationship between morphological abnormalities and genetic liability to bipolar disorder. The first study, by McDonald et al[73], investigated the association between genetic risk for bipolar disorder and schizophrenia, and tissue volume. In order to do this, a sample of both patients and unaffected relatives was collected and for each participant a measure of genetic risk (genetic liability scale) was calculated, based on their clinical status and that of their relatives.

Genetic risk for bipolar disorder was associated with grey matter deficits in the right cingulate gyrus and ventral striatum. This was true in both patients and their unaffected relatives, suggesting that the relationship was not determined only by abnormalities in patients. In white matter, genetic risk for bipolar disorder was associated with deficits in the anterior corpus callosum and bilateral frontal, left temporo-parietal and right parietal regions. As with grey matter, this relationship was present in both patients and relatives. This evidence of relationships between genetic liability and structural brain measures, present in both patients and relatives, suggests that such measures may represent endophenotypes for bipolar disorder. It should be noted however, that this study was based on a patient sample selected specifically on the basis of having psychotic symptoms and being from multiply affected families – and thus may not be representative the more broad bipolar phenotype.

Interestingly, in schizophrenia, grey matter deficits were also seen, but were spatially different from those in bipolar disorder, while white matter deficits overlapped with those in bipolar disorder. Specifically, left hemisphere white matter deficits were seen in both schizophrenia and bipolar disorder, while right hemisphere deficits were seen only in bipolar disorder. The authors concluded that their findings indicate that Kraepelin's dichotomy was neither 'wholly right nor wholly wrong'; rather, they suggest that we should think of bipolar disorder and schizophrenia as 'a sibling pair of neurogenetic syndromes'.

The second paper, by McIntosh et al[74], employed similar methodology to that of the previous study, using both VBM and a genetic liability scale. This study was also based on the same sample as two earlier VBM papers[70,72] by the same group, which looked at grey and white matter in BD patients and relatives. These groups were subdivided into those who came from families with a history of bipolar and those who can from families with a history of both bipolar and schizophrenia. These papers found that BD patients from bipolar families had reductions in grey matter density of the anterior thalamus and caudate, while BD patients from mixed families had reductions of grey mater density in the right inferior frontal gyrus and insula. Reductions of white matter density were reported in the left anterior limb of the internal capsule for BD patients from BD families. White matter density was also reduced in the right superior frontal medial gyri of unaffected relatives from mixed families. In the genetic liability study, while associations were found between genetic liability to schizophrenia and brain structure abnormalities, no such associations were detected in bipolar disorder. The finding of differences in unaffected relatives from mixed, but not BD families is interesting, the authors suggest plausibly that this may reflect a tendency for bipolar subjects from mixed families to be more 'schizophrenic like' than those from BD families. Taken together, while the three analyses support the hypothesis of structural abnormalities in bipolar disorder, unlike the study of McDonald et al[73], they do not provide evidence of a genetic basis for these abnormalities.

McIntosh et al[74] suggest a number of possible reasons for the disparity between the two genetic liability studies. While the studies employed similar methodology, it was not identical and the authors of this study point out two possible methodological

reasons for the differing results. First, in this study, the patient and unaffected relative groups also included participants from families with a mixed history of bipolar disorder and schizophrenia. The authors suggest that, while their sample may be more generalisable to other populations, classifying bipolar patients from mixed families as simply bipolar may inaccurately describe their disease phenotype. Second, the studies used different statistical methods for the VBM analysis, the former study using a permutation testing approach for testing the null hypothesis and the later using a test based on Gaussian Random Field theory and resolution element based correction. The authors suggest that, under certain circumstances, the permutation-based strategy may be more vulnerable to false positives, but acknowledge that given the inconsistent literature, the different results primarily underline the need for further investigation.

Overall, there is very little consistency in the findings from the published VBM studies. Indeed, taking all seventeen studies considered here (Case control and genetic liability), no morphological abnormality has been identified in more than four studies. The most replicated finding was of reduced grey matter volume (and or density) of the posterior cingulate, which was reported in three case control studies and the McDonald et al[73] genetic liability VBM study. However, increased volume and/or density of the same brain region was reported in three studies. With regard to white matter, as only five studies have reported investigating this with VBM, and these have not consistently replicated each other's findings, it is hard to draw conclusions at this stage. The overall lack of consistency in VBM studies reflects that from the ROI based literature and the conclusion must be the same, that larger and well-controlled studies provide the best hope of overcoming the uncertainty present in this field.

Table 1.2 VBM Studies in Bipolar Disorder

Study	Year	Sample	Areas Demonstrating Differences in Patients Relative to Healthy Controls (volume unless otherwise stated)			
			Increased Grey Matter	Decreased Grey Matter	Increased White Matter	Decreased White Matter
Lyoo et al[69]	2004	39 BD, 43 C	Density: ACC, left medial frontal, right inferior frontal, right precentral	None	Not Investigated	Not Investigated
Lochhead et al[60]	2004	11 BD, 31 C	left insula/frontoparietal operculum, ventral occipitotemporal cortex Density: left insular/frontoparietal, thalamic cortex	left ventromedial temporal, bilateral cingulate (anterior-posterior). Density: ventromedial temporal	Not Investigated	Not Investigated
Wilke et al[61]	2004	10 Adolecent BD, 52 C	basal ganglia	medial temporal lobe, OFC, ACC	Not Investigated	Not Investigated
Bruno et al[71]	2004	39 BD, 35 C	None	None	None	Density: bilateral prefrontal cortex including fronto-striatal connectiosn
Adler et al[62]	2005	32 BD, 27 C	ACC, ventral prefrontal, fusiform, primary/suplementary motor cortex (not corrected, but cluster thresholded)	superior parietal lobule (not corrected, but cluster thresholded)	Not Investigated	Not Investigated
Nugent et al[66]	2005	36BD, 65 C	None	Un-medicated BD: posterior cingulate cortex and left STG Medicated BD: left ventral PFC, left inferior frontal gyrus.	Medicated BD: left ventral PFC	Unmedicated: left medial parietal (adjacent to PCC)
Adler et al[63]	2007	33 FE BD, 33C	left thalamus, fusiform, cerebellum, mid/sup temporal gyri, paracentral lobule Density: ACC, PCC, right mid temporal, inf/sup parietal lobule, precuneus, right precentral, fusiform (not corrected, but cluster thresholded)	None	Not Investigated	Not Investigated

Study	Year	Sample	Areas Demonstrating Differences in Patients (unless stated) Relative to Healthy Controls (volume unless otherwise stated) (UR=unaffected relative)			
			Increased Grey Matter	**Decreased Grey Matter**	**Increased White Matter**	**Decreased White Matter**
Farrow et al[67]	2005	8 BD, 22 C	None	Right inf frontal / precentral gyrus, left insula, left inferior/mid temporal gyrus, left PCC. Decrease over time in ACC. (not corrected, but cluster thresholded)	Increase over time in right posterior frontal/parietal cortex, right temporo-parietal junction, left parieto-occipital junction, left parietal lobe, right cerebellum. (not corrected, but cluster thresholded)	Left Frontal Cortex. Bilateral Posterior Parieto-temporal junction. (not corrected, but cluster thresholded)
Chen et al[64]	2007	24BD 24C	Parahippocampal Gyrus	Left Middle Temporal Gyrus	Not Investigated	Not Investigated
Yatham et al[75]	2007	15 Manic FE BD, 15 C	None	None	Not Reported	Not Reported
Scherk et al[76]	2008	35 BD, 32 C	None	None	Not Investigated	Not Investigated
Almeida et al[68]	2008	27 BD, 28 C	None	Bilateral Posteromedial Rectal Gyrus, Left Parahippocampal Gyrus, Left Putamen	Not Investigated	Not Investigated
LaDouceur et al[65]	2008	20 OSB, 22 C	Left Parahippocampal Gyrus / Hippocampus	None	Not Investigated	Not Investigated
McIntosh et al[70]	2004	22 UR from BD fam 26 UR from mixed fam 26 BD from BD fam 19 BD from mixed fam	None	Grey Matter Density In patients from bipolar families: Anterior Thalamus and Caudate In patients from mixed families: Right IFG and insula density. No patient-UR differences	Not Investigated	Not Investigated
McIntosh et al[72]	2005	As Above	Not Investigated	Not Investigated	None	Patients with BD from BD fam: Left ALIC. UR from mixed: right SFG, right MFG

			Areas Demonstrating Relationship with Genetic Liability	
			Grey Matter	**White Matter**
McDonald et al[73]	2004	37 BD, 50 UR	right anterior CC, ventral striatum	anterior corpus callosum, bilateral frontal, left temporoparietal, right parietal regions
McIntosh et al[74]	2006	26 BD & 22 UR from BD family, 19 BD & 26 from mixed family.	None	None

Cellular Basis for Structural Abnormalities

The finding of white matter hyperintensities in bipolar disorder, as well as the findings of altered white matter from structural and DTI (discussed later) studies has led to the suggestion that abnormalities of myelination may play an important role in the disorder. Myelination, briefly, is the process by which glial cells form an insulating sheath around neurons. In the central nervous system, myelin is provided by oligodendrocyte glia cells. The primary role played by this insulation is to increase the speed at which electrical impulses can travel along nerve fibers. Abnormal alterations of myelin can have severe consequences and are a hallmark of neurodegenerative diseases such as multiple sclerosis.

This idea that myelination is abnormal in bipolar disorder is supported by studies at the cellular level. Uranova et al[77] looked at oligodendroglial density in the prefrontal cortex (layer 6 of Brodman area 9) and found that patients with bipolar disorder had a 29% decrease in the numerical density of oligodendroglial cells compared to controls. This finding adds to an earlier electron microscopy study by the same group that reported that patients with bipolar disorder exhibited signs of increased dystrophy, apoptosis and necrosis of oligodendroglial cells in the prefrontal cortex[78]. An earlier study by Ongur et al[79] also reported reduced numbers of glial cells (but not neurons) in subgenual Broadman area 24 (anterior cingulate). These results are concordant with other studies which show decreased levels of glial related proteins such as glial fibrillary acidic protein (GFAP)[80]. Regenold et al[81] have investigated myelin changes using staining techniques. They found that mean deep white mater (representing long range association tracts) staining intensity of the dorsolateral prefrontal cortex (DLPFC) was decreased in patients related to controls. A gene expression study by Tkachev et al[82] also provides support for abnormal glia activity, reporting that postmortem brains of bipolar patients show significant downregulation of key genes involved in myelination.

While the above studies have concentrated on glial cells and myelination, a number of post-mortem studies have also looked at abnormalities of neuronal cells themselves. Cotter et al[83] looked at both neuronal and glial size and density in the caudal orbitofrontal cortex and found that, in this area, while bipolar disorder was associated with a reduction of neuronal size (but not density) there was no difference in either glial size or density. Reduction in neuronal density has also been reported in the DLPFC[84] while reduced neuronal size and increased neuronal density have been reported in the anterior cingulate[85] and planum temporale[86] of

patients with bipolar disorder. In contrast however, a study by Bouras et al[87] reported that in the anterior cingulate, neuronal density was not increased, but decreased, along with an decrease in laminar cortical thickness (layers II, V and VI).

All of the above studies, being post-mortem in nature, cannot properly address the issue of causality. It is thus difficult to say whether the reported differences exist before the expression of the disorder, are part of the disease progression or are a side effect of other disease related processes or medication. Further, the studies examined a variety of different brain areas, with little spatial overlap, so comparing their results is not straightforward. However, despite the heterogeneity of areas investigated, the results appear to be largely consistent and the fact that complementary evidence comes from a variety of research modalities strengthens the results. Overall these studies provide strong support for the hypothesis that abnormalities of neuronal and glial architecture play an important role in bipolar disorder.

1.4.3. Functional Imaging in Bipolar Disorder.

The following section provides an overview of the most relevant functional imaging findings in bipolar disorder.

PET and SPET Studies

The first functional brain imaging studies of bipolar disorder were carried out using positron emission tomography (PET) and single proton emission tomography (SPET/SPECT). The majority of these studies investigated either 'at rest' cerebral glucose metabolism or 'at rest' cerebral blood flow rates. At rest studies investigate the activity of the cortex when no specific task is being performed. The results of these studies are summarised below. For a more detailed review of PET and SPET studies of bipolar disorder, including comparison to unipolar depression, please see Haldane and Frangou[88].

Cerebral Glucose Metabolism

In depressed bipolar patients, relative to controls, reduced cerebral glucose metabolism has been reported for the whole brain[89] and left anterior dorsolateral prefrontal cortex (DLPFC)[90]. Psychomotor-anhedonia scores of the beck depression inventory have been reported to correlate with reduced metabolism in the right insula, claustrum, caudate/putamen, and temporal cortex, and with higher metabolism in anterior cingulate[91]. Increased glucose metabolism has been reported left amygdala of depressed BD patients[92]. In the same study, amygdala glucose metabolism was also found to be elevated in a small sample (N=4) of unmedicated, remitted BD patients, but not in a small sample (N=4) of remitted patients taking mood stabilisers.

Cerebral Blood Flow

Two studies have addressed rate of cerebral blood flow (rCBF) in depressed BD patients, the first reported reduced rCBF in middle and superior frontal cortex and anterior cingulate[93] while the second reported no differences between patients and controls[94]. In manic BD patients, decreased rCBF has been reported in frontal cortex[95] and right ventral temporal lobe[96]. Increased rCBF has been reported in bilateral temporal lobes[97], as well as the anterior cingulated cortex (ACC) and left caudate of manic BD patients[98]. Finally, increased perfusion of the ACC has been reported to correlate with relapse to mania[99].

fMRI Studies

A search of the Medline database (March 2008) using the terms 'bipolar disorder' and 'fMRI' results in 453 papers. Of these, 55 are fMRI studies of bipolar disorder, the earliest of which was published in 2000 by Deborah Yurgelun-Todd's team[100]. The following is a representative overview of the published literature to date. While not exhaustive, it covers the most pertinent findings from the last eight years of functional MRI imaging in BD. For brevity, only emotional response tasks (due to their direct relevance to the disorder) and working memory tasks (due to their direct relevance to the thesis) are discussed. Only those studies that have undertaken a statistical (rather than visual) analysis of the difference between groups have been included. As the investigation of working memory tasks is central to this thesis, most weight has been given to this section.

Emotional Response tasks

Given the nature of the disorder, it is not surprising that emotional response tasks constitute the majority of all fMRI studies in bipolar disorder. The most widely used paradigm is the facial affect task, in which subjects are shown pictures of faces with emotional valance. Subjects are either specifically asked to attend to the facial affect (explicit tasks) or distracted with other instructions (implicit tasks). The results of such studies have found reasonably consistent differences (in terms of the regions, but not necessarily the direction of difference) between patients and controls; the most common of these are: (i) dysfunction of the VLPFC, both hypoactivation (BD<controls) [101, 102,103] and hyperactivation (BD>controls) [104,105], (ii) hypoactivation of the DLPFC [102,100], (iii) dysfunction of the amygdala (hypoactivation[106,105] and hyperactivation[101,102,107,104,100,105]) and (iv) dysfunction of the anterior cingulate (hypoactivation[106] and hyperactivation[101,102,108]). Less consistently, dysfunctional activation has also been reported in the, occipital lobe[101,109], posterior cingulate[106], insula[106], fusiform gyrus[110], claustrum[109], hippocampus[109], cerebellum[109], lingual gyrus[109], putamen[104], striatum[108] , orbital frontal cortex[107, 108] and nucleus accumbens[104].

Importantly, the most consistent findings (dysfunction of the VLPFC, DLPFC, amygdala and anterior cingulate) are reported in euthymic patients[101,102], indicating that dysfunction of these areas may be a trait of bipolar disorder, rather than a state effect[iv]. In order to address the effects of these local differences on activation, two recent studies have used the emotional

[iv] State effects in this case referring to effects of mood on the effect of interest

faces paradigm to investigate functional connectivity within activated brain regions. Foland et al[103] report that, compared to controls, manic BD patients demonstrate reduced regulation of the amygdala by the VLPFC, while Rich et al[111] report that euthymic BD patients demonstrate reduced left amygdala-right fusiform and left amygdala-right anterior cingulate connectivity. Further evidence for dysfunction of these areas comes from other paradigms that also attempt to tap into emotional processing. Malhi et al.[112,113], using the emotional stroop task, found that BD patients demonstrated reduced activation of left ventral PFC and right DLPFC, combined with greater activation of the left hippocampus and right amygdala. Finally, in two studies using the affective go-no-go task, BD patients demonstrated dysfunction of (among other areas) the VLPFC, medPFC and anterior cingulate.

Working Memory

The fMRI chapter in this thesis describes an investigation of the neural correlates of impaired working memory in bipolar disorder, thus working memory studies are discussed in detail. To date there have been seven published fMRI studies of working memory in Bipolar Disorder, none of which were published at the start of this thesis. These studies are presented in Table 1.3 and are discussed below. All of the studies used either the N-Back or Sternberg tasks.

N-Back Studies

Briefly, in the N-back paradigm, subjects are presented with a series of stimuli and must respond (typically via button box) when they see a stimulus that is the same as one seen 'N' presentations previously. This active condition is contrasted with a baseline condition in which subjects simply have to look for a predefined stimulus such as the letter 'X'. There are several versions of the N-back, with different stimuli and memory loads. The task is described in more detail in chapter 3.

The first fMRI study of working memory in bipolar disorder was published in 2004 by Adler et al[114], and used a single active condition (2-back), in which the patient was presented with four numeric simuli (1-4), each associated with a specific spatial location. Task related brain activity was compared between patient and control groups (each consisting of 15 patticipants). In the patient group, relative to the control group, behavioural performance was less accurate at baseline, and showed a trend to reduced accuracy at 2-back. When activation in the 2-back condition was contrasted against baseline, relative to controls, the patient group showed increased BOLD signal in: bilateral frontopolar, prefrontal and (middle and superior)

temporal cortices, anterior insula, basal ganglia, left thalamus, and left posterior parietal cortex (including lingual gyrus). Decreased signal was detected only in the posterior cingulate. The authors suggested that the increased activation seen in the bipolar group might have been due to the recruitment of an alternate processing network or cognitive strategies (or both, alternative cognitive strategies presumably resulting in different network activation) in an attempt to compensate for functional deficits elsewhere in the brain.

In the same year, Monks et al[115] published a study that used both the N-back and also the Steinberg task. Like the previous study, the N-back task used a single active condition (2-back), but this time verbal stimuli (letters) were used. 12 euthmic BD patients and 12 matched controls were scanned. No significant performance differences were detected. In patients, relative to controls, reduced task related activation was observed the anterior cingulate gyrus (extending to right medial frontal gyrus), right mid temporal gyrus, bilateral inferior frontal gyrus, mid frontal gyrus, precuneus, and cerebellum. Increased activation was seen in the left precentral gyrus, left supramarginal gyrus and right medial frontal gyrus. In a similar manner to Adler et al[114], the authors suggest that the observed differences may reflect the recruitment of 'intact slave systems' in order to support executive performance, thus maintaining performance despite potential cognitive deficits. The authors also speculate that frontal hyperactivation may reflect inefficient use of prefrontal networks.

Frangou's group[116,117] also used the N-back task, with 7 BD patients and 7 controls. This time, 3 active conditions (1,2 and 3-back, with letters as stimuli) were used, enabling the researchers to investigate the effect of varying memory load. No activation differences were seen between groups at any load level. The authors do report a group by memory load interaction, but this does not appear to have been statistically tested. Instead, for each group, the authors generated a map showing areas where activation was significantly linked to memory load; these were then compared qualitatively. The results are included in Table 1.3, but as no statistical between group tests are described, they are not discussed further[v]. The same group[118] also used the N-Back paradigm in order to investigate the effect of medication for bipolar disorder. There was however, no control group for this study and as such it is

[v] Some papers (eg Drapier et al) refer to this study as having found between group differences, however for the reasons discussed, this is not strictly true.

discussed later under *The Effect of Medication on Neuroimaging and Neuropsychological Findings.*

The most recent study is that of Drapier et al[119], which used a version of the n-back task identical to that used by Frangou et al. Drapier et al investigated three groups consisting of 20 patients, 20 relatives of patients and 20 controls respectively. The patient group was very highly selected, consisting only of BD-I patients with a history of psychosis and from multiply affected families. The study found that, during 1 and 2 back conditions, the unaffected relatives of bipolar patients had greater activation of a cluster located in the left frontal pole/ventrolateral gyrus compared to controls. A trend towards a difference in the same area was seen in the patient group during the 1-back condition. The authors concluded that hyperactivity in this prefrontal area is associated with genetic liability for the BD and represents a potential endophenotype for the disorder. Again, it was suggested that this prefrontal hyperactivity might represent inefficient prefrontal activation. While this result is interesting, there are a number of issues with the study. The first issue concerns the analysis method. The study was conducted using the XBAM software developed at the Institute of Psychiatry, London. XBAM uses ANOVA to investigate between group differences and one of the key assumptions of the standard ANOVA model is independence of observation. The study violated this assumption by including both patients and their relatives in a 3X3 ANOVA (group by task). In a standard ANOVA, each observation is treated as being independent from other observations. In repeated measures and family designs this is not true. In a repeated measures design, the multiple observations for the same subject are likely to be highly correlated, while in a family design, the shared genetics and environment are also likely to result in correlation between the scores of the members of a single family. In the case of this study, repeated measures were presumably modeled as being non-independent, but no adjustment could be made for the shared variance between family members (as the software does not currently allow for such a design). Generally it is held that violation of the independence of observation assumption increases the probability of type I (false positive) errors[120]. It is difficult however, to know how much such violations of the basic assumptions of the ANOVA model may have affected the results of this particular study. Thus, while it is important to note this limitation, one must consider that the authors' adopted a pragmatic approach that does not necessarily invalidate either the results or the conclusions of the study.

The second issue regards the patient sample, which in this study was very highly selected. It is difficult to know how far the results may generalise to a less selected sample; as all the patients had a history of psychosis, it is possible that the findings relate more to psychosis than to bipolar disorder per se. It should be noted that such selectivity, while an issue for generalisability to the broader bipolar phenotype, is also a strength of the study as it reduces sample heterogeneity and thus potentially increases power to detect differences within the specific sample chosen. Finally, it is not clear why an endophenotype would be more pronounced in relatives than in patients, although this possibly could be ascribed to normalising medication effects in the patients.

Summarising the results of the four N-back case-control studies published to date, there is relatively little consistency in the findings, with one negative and three largely non-overlapping positive studies. While Monks et al[115] and Adler et al[114] both detected group differences, the differences found, while not contradictory, do not overlap (Lagopoululos et al[121] state that these studies are in direct contrast; this is not strictly correct, as while none of the significant areas of differential activation in either study are seen in the other, given the sample sizes a null finding cannot be considered evidence of no difference). There is some overlap between Adler et al[114] and Drapier et al[119]; both studies report hyperactivation of the left frontal pole, localised in Brodman area 10 (BA10) (see Figure 1.1). This finding is compelling as, this region has been implicated in executive functions such as planning and problem solving performance[122-124], cognitive domains in which abnormalities have been reported in bipolar disorder. Hyperactivation of this area may either indicate problems with executive function or an attempt to recruit additional executive resources in order to compensate for other underlying deficits (in verbal working memory for instance).

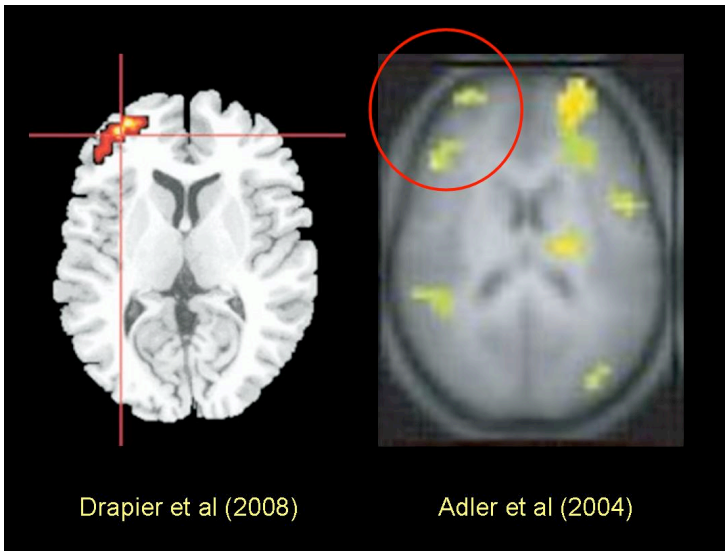

Figure 1.1 Common Finding of Frontal Pole Hyperactivation in Drapier et al & Alder et al.

Sternberg Task

In the Sternberg task, subjects first view a list of stimuli; they are then presented with a probe stimulus and have to decide whether it was in the original list. The length of the list is parametrically varied so as to investigate the effect of memory load.

Monks at al.[115] (using the same sample as for the N-Back task, above) used the Sternberg to further investigate working memory in bipolar disorder. The stimuli selected were numeric digits. In contrast the n-back, there were no significant differences between groups either in behavioural performance or brain activation. Conversely, a team from the University of New South Wales (UNSW), using another version of the Sternberg task, published two papers reporting group differences between bipolar patients and controls. Both papers were based on the same sample of 10 euthymic bipolar patients and 10 controls. The first paper[125], set out to investigate activation differences in response to implicit mood induction. Thus instead of a list of digits, a list of words with positive, negative and neutral valance was used. In this study, relative to controls, for negative and positive affect, patients showed reduced activation in anterior and posterior cingulate, medial prefrontal cortex, middle frontal and right parahippocampal gyri. For negative affect, additional differences in the same direction were found in the postcentral gyrus, inferior parietal lobule, thalamus and putamen. For positive affect, additional differences in the same direction were found in the precentral, superior temporal and lingual gyri, precuneus, cuneus, caudate, pons, midbrain and

cerebellum. The second paper[121], specifically partitioned working memory into encode, delay and response execution in order to establish how each component of working memory might be affected in bipolar disorder. In the encoding condition, relative to patients with BD, control subjects demonstrated greater activity in right IFG. In the delay condition controls had greater activation in right parahippocampal, IFG, MidFG and intraparietal sulcus, while patients demonstated greater activity in the MedFG. In the response condition, control subjects had greater activation in the anterior cingulate.

Clearly, as with the results of the N-Back studies, the results of the published studies that use the Sternberg task with bipolar disorder are not consistent. As for the N-Back results, these differences may be partially explained by the use of significantly different versions of the same task. While Monks et al. used an emotionally neutral version of the task involving numeric stimuli; the UNSW group used a word list that included words with strong negative and positive valances. It may be hypothesised therefore that the reduced activation reported in the UNSW studies is primarily due to an abnormal emotional response in the bipolar patients, rather than to any underlying cognitive deficit of working memory. The fact that neither study observed any *behavioural* between group differences strengthens this hypothesis.

Given the small number of published papers, the reasons for the inconsistency in the above working memory findings are necessarily speculation; sample size, sample selection and task variation may all be contributing factors. Sample size of particular concern in all of the above studies. According to a recent study by Thirion et al[126], sample sizes of at least 20 subjects (and preferably circa 27) are required for reliable and sensitive analyses of group differences in fMRI studies. Apart from Drapier et al[119], which had a sample size of 20, none of the published studies of working memory in bipolar disorder have group N's of more than 15. Clearly more work is needed in this area in order to clarify the existing findings.

Table 1.3 Working Memory fMRI Studies in Bipolar Disorder

Author	Year	Sample (e=euthymic)	Task	Decreased activation in BD (relative to controls)	Increased activation in BD (relative to controls)	Behavioural difference?	Notes
Drapier et al[119]	2008	20 Euthymic BD-I 20 Relatives 20 Controls	N-Back (1,2,3)	None	Relatives of BD patients had increased activation in left frontal pole/ventrolateral gyrus during 1 and 2-back. Same area showed trend towards increased activation in BD patients during the 1-back.	Patients less accurate than controls for 1,2 and 3 back. No difference between relatives and controls. Group X load interaction.	Sample highly selected for family history of bipolar disorder and personal history of psychosis.
Frangou et al.[116]	2007	7 Euthymic BD I , 7 Controls. Mood stabliser monotherapy.	N- Back (1,2,3)	None	None	No	Effect of increasing task load. Controls: localised to bilateral MF and right SFG. Patients: localised to left IPL
Haldane et al.	2007	8 BD I	N- Back (1,2,3)	NA. The study compared activation before and after lamotrigine treatment			fMRI before and after 12 weeks Lamotrigine treatment
Haldane et al	2006	Unknown	N-Back (unknown version)	'Dorsolateral PFC activation is impaired during the N-Back in remitted and moderately impaired BD patients'			This study is only available as an abstract.
Lagopoulos et al.[121]	2007	10 Euthymic BD I, 10 Controls	Sternberg	*Encode*: RIFG *Delay*: rParahippocampal, IFG, MFG, intraparietal sulcus *Response*: SFG, ACG	*Encode*: None *Delay*: MedFG *Response*: None	Patients less accurate than controls at high load.	Study designed to isolate encode, delay and response components of task.
Malhi et al.[125]	2007	10 Euthymic BD I (7 Medicated, 3 unmedicated), 10 controls	Modified Sternberg	*Negative Words*: Bilateral Postcentral FG, MFG MedFG. Right PC, parahippocampus gyrus, thalamus, Left AC, IPL. Putamen *Positive Words:* Bilateral precentral FG SFG MFG. Right AC parrahipocampal, lingual gyrus, cuneus, caudate body, pons, midbrain, cerebellum. Left PC STG, precuneus, caudate head.	None	No significant difference at any load.	
Monks et al.[115]	2004	12 Euthymic BD I 12 Controls	N-Back (2-back only)	AC, right medFG, right MTG, bilateral IFG, MFG, precuneus, and cerebellum	left precentral gyrus, left supramarginal gyrus and right medFG.	No	
			Sternberg	None	None	Patients less accurate than controls (at trend level).	
Adler et al[114]	2004	15 Euthymic BD, 15 control.	N-Back (2-back only)	Posterior Cingulate	Fronto Polar Prefrontal Cortex, Anterior Insula, Basal Ganglia, Thalamus, Temporal Cortex (middle and superior), posterior parietal cortex (including lingual gyrus)	Patients less accurate at baseline and 2-back (trend level).	

AC Anterior Cingulate PC Posterior Cingulate MFG Middle Frontal Gyrus IFG Inferior Frontal Gyrus MedFG Medial Frontal Gyrus. IPL Iferior frontal lobule

1.4.4. Diffusion Tensor Imaging in Bipolar Disorder.

A brief introduction to DTI

Diffusion tensor imaging (DTI, sometimes known as DT-MRI) is technique that uses magnetic resonance imaging to investigate the rate and directionality of diffusion of water within the brain. In order to collect diffusion weighted data, magnetic diffusion gradients[vi] are applied to the brain in a number of different directions and, for each gradient direction, a separate image is generated. If at least 6 images are collected, with gradients applied in non-colinear directions (although more directions are often used, to improve accuracy), then from these images, it is possible to calculate the diffusion tensor. This is simplest full description of diffusion, and from it, a number of scalar diffusion related parameters can be extracted for further analysis.

The two most commonly used parameters from DTI studies are the apparent diffusion coefficient (ADC) and fractional anisotropy (FA). In the brain, both of these measures vary dependent on the local tissue microstructure. ADC reflects the rate of diffusion within a given voxel; it is often reported as a mean value over all directions, and its value depends on the local tissue environment. In ventricles, which consist primarily of cerebro-spinal fluid (CSF), diffusion is relatively unconstrained, ADC is therefore high. In other areas of the brain, with higher cell density, diffusion will be restricted and ADC will be reduced.

Fractional anisotropy is a ratio that describes the relative "directionality" of diffusion at any given voxel in the brain. An FA vale of 0 indicates perfect isotropy[vii], with diffusion being the same in all directions, while an FA value of 1 represents complete anisotropy or ideal linear diffusion. Due to the random (Brownian) nature of diffusion, when unconstrained, it is

---

[vi] Here the normally homogenous magnetic field used in MRI is deliberately altered by applying a pulsed field gradient so that it varies linearly in the chosen direction. Before the gradient is applied, an 'excitation pulse' and associated gradients flip the magnetisation of the protons in the object being imaged into the transverse plane, in a process common to all MR imaging. Initially, all the protons' magnetisation vectors are aligned, and rotate (precess) in phase. As precessional frequency is proportional to magnetic field strength, however, when a field gradient is applied, the protons begin to precess at different rates, resulting in a loss of phase. After a given period of time, a gradient of the opposite sense is applied to re-phase the spins. If protons have moved (along the field gradient) during this time, re-phasing will not be perfect, which results in signal loss. From this loss of signal and other known (controlled) parameters, it is possible to infer the diffusion properties in each voxel of the brain.

[vii] Isotropy is the degree to which a property is isotropic, or invariant with respect to direction, anisotropy is therefore the degree to which a property varies dependent on direction.

equal in all directions. FA is dependent on both the degree and the type on constraint present, if the constraint is equal in all directions ADC will be reduced, but FA will be relatively unaffected. If however, diffusion is only restricted in certain directions, FA will be increased. Areas of the brain that consist mainly of large bundles of white matter tracts have relatively high FA values; this is because the structure of the local tissue is such that while diffusion is constrained across the tracts (due to oriented structures including the axonal cell bodies themselves and their myelin sheaths), it is still relatively free parallel to the tracts. In areas of grey matter, FA is greater than in CSF, but is less than in white matter.

Figure 1.2 demonstrates the effects of constraints upon diffusion. The purple lines represent the movement of a single water molecule, while the red outlines show the overall diffusion rate and direction. Compare A to B: water is diffusing more slowly in B than in A, but the diffusion is isotropic in both cases. Thus, in both A and B, FA is 0, but ADC is lower in B than in A. Now compare B to C: in A, diffusion is isotropic, but in C, diffusion is faster along the Y-axis than the X and Z axes. Thus, in this comparison, mean ADC is similar in both balls, but FA is much greater in C than in A.

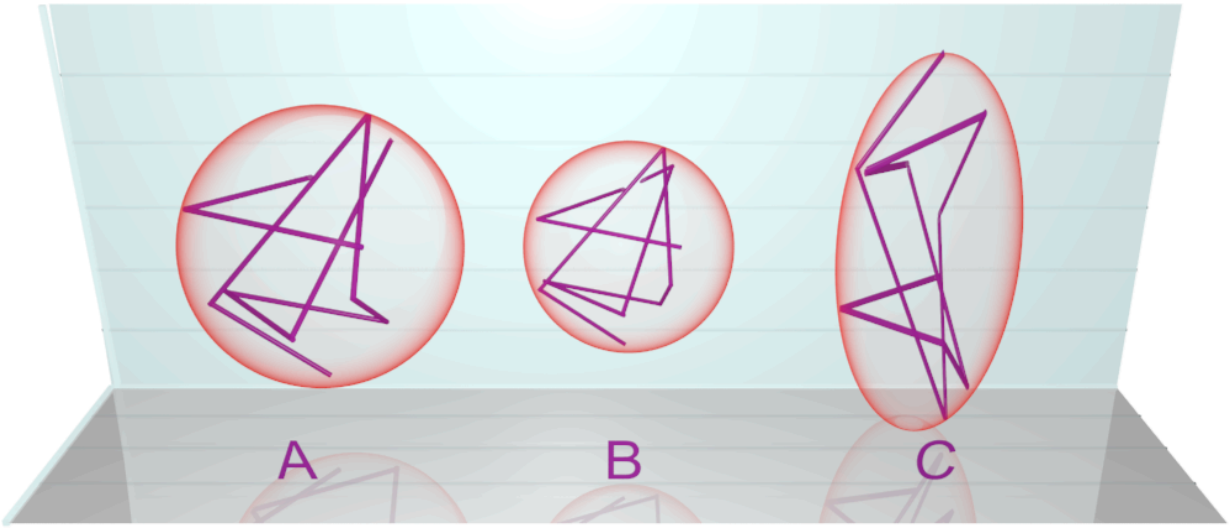

Figure 1.2 FA and MD.

(Adapted from an image on Chris Rorden's online DTI analysis page, http://www.sph.sc.edu/comd/rorden/workshop/fsl/dti/)

It is worth noting at this point that interpretation of differences in FA and/or ADC is extremely difficult. In a white matter tract, reduced FA might represent a reduced number of axons, but equally it might represent reduced myelination, reduced directional coherence,

lesions or an increased number of 'u-shaped' fibres or other axons crossing the tract of interest. Thus any differences must be interpreted with due regard to the locations in which they are found and to our knowledge of the normal anatomy of these areas.

Mean Diffusivity (MD), Apparent Diffusion Coefficent (ADC) and Trace.

ADC, MD and trace are very similar concepts and are often used interchangeably, which can be confusing. An ADC value is an indication of the rate of diffusion in a particular direction, and may include a subscript to indicate the measurement direction (e.g ADC_x; ADC_z). MD is the average of ADC measurements made in three mutually orthogonal directions. *Trace* is a mathematical term which in this context refers to the sum of the diagonal elements of the diffusion tensor, which is in turn equal to the sum of the diffusivities measured in any three mutually orthogonal directions; thus, trace = MD * 3. As a result, for most purposes MD and trace, being directly proportional, can be considered equivalent. Unlike MD and trace, ADC may refer to a measurement of diffusion in just one direction (sometimes, when full DTI information is not required, in order to reduce scanning time, diffusion data is gathered in only one direction). *Mean* ADC, which is ADC averaged across directions, is equivalent to MD, although confusingly the term ADC alone (without either a directional subscript, of the use of the word mean) is often also used in this context.

Tractography.

Beyond FA and MD values, DTI imaging can also provide information about the principal direction of diffusion for each voxel. As previously discussed, in white matter tracts, water tends to diffuse fastest in a direction parallel to the tract. It is thus possible to use information about the principal direction of diffusion in order to "virtually reconstruct" the paths of white matter tracts within the brain – a technique known as tractography. Such virtual reconstruction relies on the assumption that if two adjacent voxels have a similar principal direction (and point towards each other) they are more likely to share underlying cellular architecture than two adjacent voxels with divergent principal directions. It follows that the more similar the principal diffusivity direction of two adjacent voxels, the more likely they are to belong to the same white matter tract. Essentially, to reconstruct a tract, one selects an anatomically plausible 'seed' voxel and then iteratively calculates which other voxels may belong to the same tract. This is of course, a simplification, for more detailed explanations of tractography methodology please see Basser et al[127] and Jones[128].

Studies to Date

To date there have been thirteen studies of bipolar disorder using DTI methods - none of which had been published when I began the studies towards this Thesis. Seven of these studies have used region of interest (ROI) methodology, two have used both whole brain voxel based morphometry (VBM) and ROIs, one has used just VBM, two have used tractography methodology and one has used tract-based skeletal statistics. The results of these studies are summarised in Table 1.4 and are discussed below.

The earliest published study is from Adler et al[129]; this study investigated DTI measures in a group of 9 BD patients and 9 matched controls. ROIs were chosen based on the authors' previous fMRI findings of differential brain activation in patients. FA and ADC values were extracted from four areas that were 15, 20, 25 and 30 mm superior to the anterior commissure. FA values were found to be significantly lower in BD for two regions, those 15 and 25mm superior to the anterior commissure. No other differences were found. The authors suggest that reduced FA without significant ADC change may reflect loss of bundle coherence in the absence of other cell damage that would affect ADC. It was also noted however, that the sample size might have been too small to detect differences in ADC (and FA in the two regions with no difference). White matter in the affected areas links the prefrontal cortex with both subcortical and other cortical regions; the authors therefore speculated that their findings may represent evidence of a disruption in network connectivity.

In 2005, Haznedar et al[130] published a study comparing a group of 33 BD patients (11 BD-I, 6 BD-II and 16 cyclothymic) with 34 matched controls. ROIs were placed bilaterally in 6 regions of the internal capsule (3 anterior, 3 posterior) and 5 regions of the frontal cortex (2 anterior frontal lobe white matter, 1 superior longitudinal fasciculus (SLF) and 2 anterior fronto-occipital fasciculus). BD patients differed from controls in the posterior, but not anterior internal capsule, with patients subjects having lower FA values bilaterally; these differences were significant in the BD-I and cyclothymia groups, but not in the BD II group. BD patients also demonstrated differences in frontal white matter, having lower FA in the anterior frontal-occipital fasciculus and higher FA in anterior frontal lobe white matter; post-hoc analyses found that these differences were significant in the BD-I group only. It should be noted that in this study the cyclothymic patients were all recruited from a population of pathological gamblers and thus may not be representative of the larger population. The authors tentatively concluded that, in the frontal cortex ROIs, given the proportion of local

intergyral fibres, the finding of higher anterior frontal lobe white matter FA in BD may represent a reduction in the proportion of low FA 'u-shaped' short intergyral fibres and an increase in fibres with a vertical orientation and/or axial longitudinal fibres with high FA. Conversely, the decreased FA in the fronto-occipital fasciculus presumably represents a decrease in the number or coherence of fibres. With regard to the finding of reduced FA in the posterior internal capsule, the authors hypothesised a non-specific disorganisation of fibres in BD and suggest that their data warrant further investigation of thalamo-cortical connections.

Also in 2005, Beyer et al[131] published a DTI study of 14 BD patients and 21 matched controls. In this study, ROIs were placed bilaterally in white matter tracts of the orbital frontal cortex and DLPFC as well as in the superior and middle frontal gyri. No FA differences were found, but the BD group was found to have bilaterally increased ADC in the orbitofrontal cortex. The authors speculated that the observed differences in ADC, in the absence of FA differences might represent increased intracellular water, without changes in myelination or axonal disruption. However, they also suggest that the study may simply have been underpowered to detect FA changes. It was noted further that the prefrontal ROIs were close to those chosen by Adler et al, yet, in contrast Adler's study, showed no differences in FA. Difference in exact ROIs, methodology, samples, and sample size are suggested as reasons for the disparity.

In 2006, Regenold et al[132] published a pilot study comparing eight BD-I patients (all with history of psychosis) and eight controls. While this study used ROI methodology, it was global in that an average of 40 ROIs were chosen for each subject, covering the frontal, temporal, parietal and occipital lobes. Due to the small size of the sample and the large number of ROIs, the researchers chose to first compare the mean (mean) ADC of all ROIs and then to perform additional regional analyses. The group mean ADC was significantly different, with the BD patients having increased mean ADC. Similar differences were also identified at a regional level, but were not significant following Bonferroni correction for multiple comparisons. In this study, increased global white matter ADC in BD has been interpreted, reasonably, as being indicative of white matter disease. The authors note that as their BD sample was one of chronic, severely ill, treatment resistant patients with psychotic symptomatology, this might account for the differences found in this, but not other studies. A serious confound in this study is that control subjects were recruited from a group of

patients who had undergone neurological assessment for possible stroke. Even though none of the controls had abnormalities that could be detected by a neurologist examining the DWI images, they clearly represent an unusual control group.

Yurgelun-Todd el al.[133] compared a group of 11 BD-I patients and 10 controls. ROIs were placed bilaterally in the genu of the corpus callosum, two bilateral regions of forward projecting white matter anterior to the anterior cingulate and in the midline of the splenium. Relative to controls, patients were found to have significantly higher FA in the midline of the genu. The authors also report that, while controls showed significantly higher FA of the splenium relative to the genu, this was not seen in the BD group. It should be noted that these two results are probably different manifestations of the same difference. It is also not clear from the paper if any attempt was made to control for multiple comparisons and thus these results must be treated with caution. The prefrontal regions investigated in this study were close to those reported by Adler et al[129] and Beyer et al[131], the results of this study being in agreement with Beyer's finding of no difference and in opposition to Adler's finding of reduced FA in patients.

All of the studies reviewed so far share the same major potential confound, that of the effects of medication (combined with small and poorly selected samples). In order to address this issue, Alder et al[134] conducted a study of 11 medication naive adolescent BD patients and 17 controls. ROIs were placed bilaterally in superior, middle and inferior frontal white matter as well as in superior, middle and inferior posterior white matter. In the BD group FA was found to be significantly reduced in the superior fontal cortex; this finding was bilateral, but more significant in the left than right hemisphere. No other differences in FA or ADC were detected. As in their previous (2004) study, the authors concluded that the evidence pointed towards a loss of coherence, but not cellular damage of frontal white matter.

Unlike the previous studies, Frazier et al.[135] used a whole brain analysis methods, as well as ROI methodology to investigate potential DTI differences in BD patients. In this study, they investigated three groups, a group of 10 children with BD, a group of 7 children at risk for BD and a group of 8 controls; the mean ages for the groups were 9.2, 9.2 and 8.9 years respectively. Although the study was essentially a whole brain, voxel based analysis, ROIs were defined prior to the analysis and diffusion measures were extracted in these areas from the spatially smoothed and normalised data from the voxel based analysis. These ROIs were located bilaterally in the superior longitudinal fasciculus, frontal orbital white matter and the

cingulated-paracingulate (defined as tracts to/from the anterior and posterior cingulate and paracingulate gyri). More specific details as to the exact locations are not available. In the ROI analysis, compared to controls, the BD group had significantly reduced FA in bilaterally in the SLF and the cingulate-paracingulate white matter. The 'at risk' group had also had significantly lower FA bilaterally in the SLF, but not in the cingulate-paracingulate. The whole brain analysis revealed additional group differences in the left OFC and the right corpus callosum body, with the BD group (but not the 'at risk' group) having lower FA in all areas. A major concern in the interpretation of this study is the sample being studied, but in terms of size and selection criteria. While it is obviously of interest to study the earliest stages of BD, the issue of pre-pubescent BD remains controversial. Indeed it is not clear whether the 'disorder' represented by this sample is an early (and by its early onset, perhaps more severe) stage of adult BD, a distinct subtype of BD, or something else altogether[136]. Assuming that this sample does represent an early onset of BD, the results would appear to support Adler's finding of reduced frontal FA. In addition, the finding of reduced FA of the SLF in the 'at risk' group suggest that reduced FA of the SLF may represent an endophenotype of BD, a finding that certainly merits further investigation.

Basing their work on the results of previous DTI and structural studies, Houenou et al.[137] used (in 16 BD patients and 16 controls) a tractography method to investigate white matter tracts in BD. In particular, the authors were interested in investigating the connections between the temporal lobe and the ventral PFC and OFC, areas that are primarily connected via the uncinate fasciculus (UC). To investigate such connections, for each subject, seed masks were placed bilaterally in the subgenual cingulate gyrus (SG) and the amygdalo-hippocampal (AH) complex. A tractography algorithm was the used to reconstruct fibres or 'streamlines'[viii] connecting the SG and AH seeds. Mean FA and mean ADC values were also calculated for reconstructed fibre tracts. The number of streamlines connecting the left SG and left AH were found to be significantly higher for the BD group than the for the control group. No differences were found in the right hemisphere. In the BD group, but not controls, a left-right asymmetry in the number of streamlines was found (this presumably reflects the group difference in left sided SC-AH streamlines). There were no differences in FA or ADC either within the reconstructed fibres or the seed-masks. The authors suggest that these

[viii] The concept of 'streamlines' is borrowed hydrodynamics, along with methodology for streamline reconstruction.

results may represent an increase in the number, but not density, of axons within the tract studied. Further, they suggest that this may shed some light on the mechanisms of the observed dysfunction of limbic structures in BD (discussed earlier with regard to functional imaging). However, as with all DTI data, it is important to note that this interpretation of tractography results is controversial. Differences in the number of mathematically reconstructed streamlines cannot be directly interpreted in terms of either increased numbers or density of axons.

2008 has seen the publication of five papers in bipolar-DTI, all of which have been conducted with relatively large sample sizes. The first of these recent publications is that of Bruno et al[138], which is the second published DTI study in bipolar disorder to use VBM. The study consisted of 36 patients (25 BD-I, 11 BD-II) and 28 matched controls. FA was found to be significantly lower in the patient group (relative to controls) in an area located at the junction of the middle and inferior temporal gyri, which corresponds to the location of part of the inferior longitudinal fasciculus. MD was found to be significantly higher in the patient group (relative to controls) in bilateral prefrontal white matter, including the anterior part of the frontal-occipital fasciculus. Increased MD was also found in an area of right posterior frontal white matter, including the posterior segment of the frontal-occipital fasciculus and the posterior section of the body of the corpus callosum.

Wang et al published two DTI-bipolar papers. The first paper[139] investigated the FA of the corpus callosum (in a sample of 33 bipolar patients and 40 controls) with both ROI and whole brain voxel based methods. The paper is not clear about whether the voxel based method was restricted to the CC. Nevertheless, both methods found evidence of reduced FA in the CC of patients with bipolar disorder. For the ROI analysis, changes were localised to the anterior and middle CC, while in the voxel based analysis charges were similarly localised to the genu, rostral body and anterior midbody. In their second paper[140], in a (presumably overlapping) sample of 42 patients with bipolar disorder and 42 controls, Wang et al used ROI methodology to investigate FA of the cingulum. The study found resulted FA in the anterior, but not posterior cingulum.

Versace et al, in a study of 31 BD-I patients and 25 controls, used a very new technique, tract based skeletal-statistics (TBSS) to investigate FA differences at the whole brain level. TBSS offers a novel way of registering white matter images, by registering all subjects to a standardised white matter skeleton[141,142]; this method may overcome some the registration

and smoothing problems common to voxel-based methods. The study reported increased FA (in patients relative to controls) of the left uncinate fasciculus (UF), left optic radiation and right anterothalamic radiation. Reduced FA was found in the right UF only.

Finally, the most recent publication, by McIntosh et al[143], used tractography to investigate FA differences in a sample of 40 BD patients and 49 controls, as well as 25 patients with schizophrenia. The study examined two white matter tracts, the UF and the anterior thalamic radiation. Decreased FA was found in bipolar and schizophrenic patients for both tracts. The authors concluded that this may reflect a common aetiology for both disorders. This study is in striking contrast to that of Versace et al, which reported almost diametrically opposite results.

As is often the case in MRI studies, it is difficult to directly compare the results of the various studies looking at DTI in bipolar disorder; this is primarily due to the different methodologies, samples, and (perhaps most problematic) the selection of ROIs in each study. Nine of the thirteen studies found evidence of reduced FA in bipolar disorder, while only three found evidence of increased FA. In the 7 studies that also investigated MD (or its analogues), 3 found increased MD in bipolar patients, 4 found no difference and none found a reduction in MD. Due to the diverse locations of the regions studies, it is not easy to comment firmly on the spatial overlap between studies. Taking the studies together, the findings remain contradictory, including: abnormality of uncinate fasciculus FA, both increased[144] and decreased[144,143]; abnormality of CC FA, both increased[133] and decreased[135,139]; abnormality of anterior thalamic radiation FA, both increased[144] and decreased[143]; abnormalities of frontal white matter FA (covering a large area of WM), both increased[129,131,134,135] and decreased[130,138]; reduced FA of cingulum/cingulate WM[135,140] and reduced FA of temporal white matter[138]. Clearly then, there is no consensus yet on white matter abnormalities in bipolar disorder. Overall however, the strongest evidence does appear to indicate that white matter is abnormal in bipolar disorder, and this mostly manifests as reduced FA.

Until recently, the majority of the published studies had very small sample sizes. With such small samples, one cannot say with any certainty whether the findings may generalise to the broader patient population; indeed the early studies would now be considered un-publishable. Additionally, the some of the studies had somewhat loose sampling criteria, one combined BD-I and BD-II patients, while the other combined BD-I, BD-II and cyclothymic compulsive

gamblers. While the former may be seen as pragmatic, the later is less easy to justify. However, more recent studies have used larger and better-defined samples. This trend is encouraging and should result in more robust and replicable findings in future (in contrast to the structural imaging literature, in which sample sizes remain too small, despite many publications[59]).

Table 1.4 DTI Studies in Bipolar Disorder

Author	Year	Sample (e=euthymic)	Study Type/ Parameters investigated	ROI Areas Tested	Areas of decreased FA / ADC/ MD, in patients relative to controls	Areas of increased FA / ADC / MD, in patients relative to controls	Other Comparisons/Notes
Mcintosh et al	2008	40 BD, 49 control	Tractography Tract based FA averages.	NA	Decreased FA of uncinate fasciculus and anterior thalamic radiation.	None	Study also investigated 25 patients with schizophrenia, finding FA deficits in the same areas.
Versace et al	2008	31 BD-I, 25 control	TBSS	Uncinate fasciculus, Anterior thalamic radiation	Reduced FA of: Right uncinate fasciculus (UF).	Increased FA of: left uncinate, left optic radiation, right antero-thalamic radiation.	Correlation found between age and FA in patients in bilateral UF and right antero-thalami radiation.
Wang et al B	2008	42 BD, 42 control	ROI	Cingulum, anterior and posteiror	Reduced FA of anterior, but not posterior cingulum		
Wang et al A	2008	33 BD, 40 control	ROI Whole Brain	Corpus Callosum	VOI: Reduced FA in anterior and middle CC Whole Brain: Reduced FA in genu, rostral body and anterior midbody of CC		
Bruno et al.	2008	36 BD (25 BD-I, 11 BD-II), 28 controls	Whole Brain FA and MD	NA	FA reduced in ITG and MidTG gyri (incorporating ILF)	MD increased in right posterior frontal and bilateral prefrontal white matter.	Areas of increased MD overlap with those showing WM density decrease in same subjects
Frazier et al.	2007	10 BD children, 7 at-risk for BD children, 8 controls	Whole Brain & ROI FA	SLF, Frontal Orbital WM, cingulated-paracingulate.	**ROI:** BD children had reduced FA in bilateral SLF and cingulate-paracingulate. At risk children had reduced FA in bilateral SLF. **Whole Brain**: BD children had reduced FA in Left OFC, Right CC Body.	None	FA: BD < at-risk for BD, in cingulate-paracingulate
Yurgelun-Todd DA et al	2007	11 BD-I (e), 10 controls	ROI FA + MD	Midline and forward projections of genu. midline of splenium.	None	Increased FA in midline of genu.	Region X Group Differences. FA of genu < splenium in controls but not patients. Significant +ive correlation between MD and age at onset in splenium
Houenou et al	2007	16 BD (e), 16 controls	Tractography Tract based FA & ADC averages.	Tracts passing through subgenual cingulate to amygdalo-hippocampal complex.	None	None	Significantly more reconstructed fibres per seed mask volume in BD, compared to controls. No FA / ADC differences in tracts.

Continued Overleaf

Author	Year	Sample (e=euthymic)	Study Type/ Parameters investigated	ROI Areas Tested	Areas of decreased FA / ADC/ MD, in patients relative to controls	Areas of increased FA / ADC / MD, in patients relative to controls	Other Comparisons/Notes
Regenold et al	2006	8 BP, 8 controls	ROI ADC only	approx 40 ROIs. Frontal, temporal, parietal, occipital lobes	None	ADC: group mean: BP > control	
Adler et al	2006	11 First Episode adolescents, 17 controls	ROI FA & Trace	Bilateral superior, mid & inferior frontal. Bilateral superior, mid & inferior posterior	Reduced FA in superior frontal cortex, strongest on left side.	None	
Haznedar et al	2005	33 (11 BD I 6 BD II, 16 cyclothymia), 34 Controls	ROI FA	3 Regions of anterior IC, 3 posterior IC, 5 FC	**Combined BD group:** Decreased FA in posterior IC, anterior frontal occipital fasciculus. **Subgroups:** BD-I only showed reductions.	**Combined BD group:** Increased FA in Anterior Frontal White Matter. **Subgroups:** BD-I only showed higher FA in anterior frontal WM and SLF.	**Combined BD group:** relative to controls, patients showed decreased anterior posterior FA asymmetry in anterior genu, as well as loss of left right asymmetry.
Beyer JT et al	2005	14 BD, 21 Control	ROI FA + ADC	Bilateral: OFC superior and middle frontal gyri	Nothing significant, but lower FA reported for all ROIs.	Increased ADC in bilateral OFC	
Adler et al	2004	9 BD, 9 Controls	ROI FA + ADC	15, 20, 25 and 30 mm superior to AC	FA decreased in ROIs 15 and 25mm superior to AC. No ADC changes	None	

Abbreviations: CC: corpus callosum, IC: internal capsule, ILF: inferior longitudinal fasciculus, ITG: inferior temporal gyrus, MidTG: middle temporal gyrus, OFC: orbital frontal cortex, ROI: region of interest, SLF: superior longitudinal fasciculus, TBSS: tract-based skeletal statistics,

1.4.5. Confounds in Bipolar Disorder Reseach: Medication and Mood Symptoms.

Studies of patients with bipolar disorder must deal with two particular confounds, mood symptoms and medication. Both are discussed below.

Mood Symptoms

It is well established that mood symptoms can affect cognitive performance and it is equally possible that this may affect other measures such as functional activation in fMRI studies. Mood thus clearly presents a potential confound in investigations of differences between bipolar patients and controls. One way to address this problem is to select a subset of bipolar patients who are currently in remission ('euthymic' patients). However, the definition of bipolar euthymia differs from study to study; these definitions are also somewhat arbitrary and do not exclude the possibility that patients may still have subtle mood abnormalities that do not meet full criteria for a manic, hypomanic, depressive or mixed episode. Importantly, a recent study by Henry et al[28] has shown that euthymic patients have higher affective lability and more intense emotions than controls[28].

Given the above, it is quite possible that sub-syndromal variation in mood symptoms may be responsible for many of the cognitive differences reported in case-control studies. This is especially important given the relatively subtle nature of some of the differences reported to date (e.g. cognitive deficits). Although a number of studies[29-31] have attempted to 'control' statistically (i.e. co-vary) for sub-threshold mood symptoms; despite its common use in the psychological literature, the validity of this statistic approach is under debate[9]. In theory, a very strict definition of euthymia that excludes sub-syndromal mood symptoms would overcome this caveat, but would be logistically difficult.

Medication

From clinical observation it is clear that the psychoactive medications used to treat patients can have a significant effect on their cognitive function. Indeed, in bipolar disorder, such

---

[9] While adjusting for the effect of covariates is a valid and useful way to specify a model and remove noise from data, it can only be done given certain conditions. If groups are significantly different on a variable due to their group membership, it is not valid to adjust for that variable. This is because, in such cases, it is not possible to partition the variance due to the group membership and that due to the covariate. This is exactly the case when comparing patients with sub-syndromal symptoms to healthy controls. Please see 2.11.

medications are specifically prescribed in order to alter abnormal cognitions and emotions. However, the issue of medication as a confound in studies of bipolar disorder is so difficult to address that researchers often either just ignore it, or dismiss it as a minor confound. In the following section, I describe the existing the exiting evidence regarding the effect of medication on cognition and neuroimaging.

Animal Models

Neuro-protective effects and alteration of neuronal morphology resulting from treatment Lithium and Valproic acid have been demonstrated in a number of pre-clinical studies (see Shaltiel et al[145] and Chen and Manji[146] for reviews). A recent study of simulated stroke (90 minute middle cerebral artery occlusion) in the rat brain found that chronic administration of lithium resulted in recovery of BOLD response following the simulate stroke, although it was not clear what the underlying mechanism was[147]. This has specific implications for functional MRI studies, which measure the BOLD response.

Human Studies

Neuropsychological effects of medication for bipolar disorder.

A few published studies have specifically addressed the effects of mood stabilisers on cognition in bipolar disorder. One did not find any significant differences between medicated and unmedicated patients[148], but also failed to find patient-control differences. A more recent study[149] estimated the impact of anticonvulsants and lithium treatment on cognition. The study did not have a normal control group, instead using a normative database. The authors concluded that each treatment was associated with differing degrees of 'cognitive toxicity', with lamotrigine and oxcarbazepine having the least impact on cognition, lithium having intermediate impact and topiramate, valproic acid carbamazepine having the most impact. There is also an often overlooked literature on the effects of medication in normal controls. Lithium treated volunteers have been shown to exhibit deficits on a variety of cognitive and psycho-physiological measures including decreased vigilance[150,151], semantic reasoning[152], memory and learning[153,154]. Neuroleptic medication has also been associated with cognitive deficits in normal controls in a similar variety of domains[155]. It is possible that the deficits seen in acutely medicated healthy volunteers may remit over time, and are therefore less likely to represent a confound in chronically mediated patients, but this is largely untested. If medication does, as seems likely, have a significant effect on cognitive function, then this has obvious consequences for functional and (perhaps) structural imaging results as well.

Structural effects of medication for bipolar disorder.

A few studies have investigated the effect of medication on brain structure. One study of amygdala and hippocampus volume in bipolar patients treated with and without lithium, showed that those patients treated with lithium exhibited significantly increased amygdala and hippocampus volumes compared to those not treated with lithium[156]. A further study[157] of hippocampal volume compared three groups of patients with BD and a control group. The groups consisted of (i) 12 patients treated with a short course (1-8) weeks of lithium, (ii) 7 patients treated with 1-8 weeks of valproic acid or lamotrigine, (iii) 9 patients who were either unmedicated or with <5 days of medication and (iv) 30 controls. The study found that the hippocampal volumes of lithium treated patients were significantly higher than those of unmedicated patients and controls (which did not differ). Bearden et al[158] report significantly greater grey matter density of the right anterior cingulate in lithium treated (n=20) compared to non-lithium treated patients (n=8). Finally, in their meta-analysis, Kempton et al[59] carried out a meta-regression testing for the effects of lithium and reported that lithium was associated with an increased in grey matter volume. Overall, although the evidence is sparse, it does suggest that medication may have significant effects on brain structure; this in turn may have consequences for the interpretation of imaging data from medicated patients.

Functional effects of medication for bipolar disorder.

Relatively little work has been done to explore the effects of mood stabilisers on brain activation. Only one fMRI study has addressed this issue in a non-psychiatric population[159]. The study investigated the effects of mood stabilisers by treating (for 14 days) healthy controls with either: sodium valproate (n=12), lithium (n=9) or placebo (n=12). The volunteers were scanned twice (once at baseline, once post-treatment) with three functional paradigms. A significant group effect was found for all three tasks. While the effects of medication were task and region dependent, both lithium and sodium valproate treatments were associated with decreased BOLD signal.

A number of studies have looked at the effects of medication in patient populations. Caligiuri et al[160] compared the functional activations of bipolar patients taking antipsychotic and mood stabilising medications during a manual reaction time task. Patients taking medication demonstrated significantly reduced activation relative to unmedicated patients. The authors suggest that this may represent a normalising effect of medication. More recently, Haldane et al[118] investigated the effects of lamotrigine treatment in two functional

tasks (the n-back and an angry facial affect recognition task). In this study 8 previously medicated patients were switched to lamotrigine and scanned twice, once at baseline and once at study end. For both tasks, following treatment, patients demonstrated increased activation in (primarily) the prefrontal cortex and cingulate gyrus. Unfortunately, in this study there was no control group, so it is not possible to rule out that this was related to practice/learning effects. The same group has also published a similar facial affect study (but with sad faces), that appears to be from the same dataset, this time with an added control group. However, as the control group was only scanned once, the aforementioned problem remains[161]. Nevertheless, the authors concluded that lamotrigine treatment resulted in a normalisation of altered brain function in bipolar disorder.

Other methodologies have been used to investigate the effects of mood stabilisers (although this has been based around their use as anticonvulsants). In particular, two studies[162,163] have used positron emission tomography (PET) to investigate the effect of treatment with valproic acid in healthy volunteers. Both studies found that valoproic acid reduced glucose metabolism. The second of these studies[162] also investigated cerebral blood flow (CBF) and found this to be significantly reduced by treatment with valoproic acid. As both CBF and glucose metabolism are closely linked to the strength of the BOLD signal, these medication effects must be considered when assessing reports of reduced function activation in patients with bipolar disorder. It has also been reported (using SPECT) that lithium withdrawal results in increased posterior cortex perfusion but decreased anterior cingulate perfusion.[99]

Using Family and Twin Studies to Address Confounds of Medication and Mood Symptoms

As discussed above, medication and sub syndromal symptoms are important confounds that are difficult to address in case control studies. The effects of medication and sub-syndromal mood symptoms on cognitive function and its neural correlates are far from clear, a fact that undoubtedly stems from the difficulty of conducting studies that control for such factors.

Family and twin studies are (as discussed earlier) important in the confirmation of the endophenotypic status of abnormalities observed in patients. An extremely important, and often understated, ancillary function of such studies is that they enable us to investigate such endophenotypes without the confounds of medication and (to some extent at least) sub-syndromal mood symptoms. This is because the unaffected relatives of bipolar patients are

much less likely than the patients to be taking medication or to exhibit prominent mood symptoms. If these two assumptions prove correct in a group of unaffected relatives, then it is reasonable to conclude that any abnormality present in both patients and their relatives is due to the shared familial (genetic and environmental) influences, rather than medication or mood symptoms.

An alternative approach is to study medication naïve patients. However, this is logistically difficult as the mean time from onset of bipolar disorder to treatment is 15 years, and early intervention services are primarily focused on schizophrenia, not bipolar disorder. One could also study a sample of patients who are medication free, but such a sample are unlikely to be representative of bipolar disorder more generally.

1.5. Specific hypotheses:

The specific hypotheses to be tested in this thesis are:

Verbal Working Memory

1. Relative to controls, bipolar patients will demonstrate quantitatively different brain activation in response to a verbal working memory task.

2. Qualitatively similar, but quantitatively less severe differences in brain activation observed in bipolar patients will also be seen in their unaffected co-twins.

3. These intra-twin pair differences will be more pronounced in fraternal than in identical twins, reflecting the genetic basis of these differences.

Diffusion Tensor Imaging

4. Relative to controls bipolar patients will demonstrate abnormalities of white matter integrity. These differences will most likely be located in frontal white matter and/or long range association fibres.

5. Qualitatively similar, but quantitatively less severe abnormalities will be present in their unaffected co-twins.

6. These intra-twin pair differences will be more pronounced in fraternal than in identical twins, reflecting the genetic basis of such differences.

2. Methods

This chapter outlines the primary methods that are common to chapters 3-5 (more specific methods are discussed in detail in the relevant chapters).

2.1. Ethical approval for the Maudsley Bipolar Twin Study

Ethical Approval for the study was granted by the NHS South East Multi-Centre Ethics Committee (#059/99).

2.2. Naming Conventions Used in this Thesis

For the purposes of brevity and legibility, the following naming conventions have been maintained.

- Twins with a diagnosis of bipolar disorder (I or II) will be referred to as 'BD twins'.
- Cotwins of twins with a diagnosis of bipolar disorder, who themselves do not have a diagnosis of bipolar disorder; will be referred to as 'unaffected cotwins'.
- Where the above twins come specifically from pairs discordant (where only one twin has a diagnosis) for bipolar disorder, the twin with a diagnosis will be known as discordant bipolar (DB), while the twin without a diagnosis will be known a discordant non-bipolar (DNB)
- Control twins will be referred to simply as 'control twins'.

2.3. Recruitment of Subjects.

The target population were identical and non-identical twin pairs, where either one or both twins had a current DSM-IV diagnosis of Bipolar disorder (I or II). In the general population, the prevalence of bipolar disorder is estimated at about 1%, while the prevalence of twins in the general population is also about 1%. The prevalence of twins with diagnosis of bipolar disorder can therefore be estimated at about 1 per 100,000. Given such a rare population, it was necessary to recruit using any practical means. Recruitment was nationwide (United Kingdom) and the primary recruitment methods were:

- **Contact with health professionals:**

  Psychiatrists and other health professionals were contacted via letters, advertising and word of mouth to make them aware of the study and were asked to pass on information to any potential subjects.

- **Advertising:**

  Adverts were placed in national and local newspapers as well as in specific

user group publications such as *Pendulum*, the Manic Depression Fellowship's quarterly newsletter. Flyers for the study were also distributed in hospitals, clinics and chemists. Links were also placed on various internet sites such as Wikipedia.org and self-help groups.

- **Talks:**
  Researchers from the Maudsley Twin Studies and other affiliated studies talked at meetings of the Manic Depression Fellowship and Rethink as well as at mental health conferences such as the Stanley Bipolar Disorder conferences.

Control subjects were identical and non identical twin pairs, without a significant personal or family psychiatric history (see below). Control subjects were recruited primarily though advertisements in the local press (the London Evening Standard and the London Metro). Controls were chosen to match (at group level) the patient group on age, gender, ethnicity and parental socio-economic class.

2.3.1. Note on Methodological Considerations Over Recruitment Strategy:

The wide range of recruitment methods and the nationwide nature of recruitment meant that the sample was relatively heterogeneous and free from potential biases associated with recruitment from one area or clinical trust. However, as many of the volunteers joined the study in response to adverts (approximately 70% of the bipolar twins, all of the controls) this may have introduced a self-selection bias. A self-selected sample may have different clinical characteristics than a sample primarily recruited from psychiatric services. Specifically, a self selected sample may include a wider range of volunteers, including those with little current contact with psychiatric services, who may therefore be less severely affected by their psychiatric condition. Samples ascertained only from psychiatric services, by contrast, are likely by their nature, to be largely restricted to volunteers with more severe psychiatric issues. It is possible therefore, that by encompassing a wider range of the bipolar spectrum; a largely self selected sample may not be representative of people in regular psychiatric contact. However, by the same argument, a self selected sample may be more representative of bipolar disorder within the general population.

For the overall bipolar sample the mean number of reported manic (or hypomanic for BD II) episodes was 7.9 (s.d. 16.9, range 1-100), while the mean number of

depressive episodes was 7.8 (s.d. 11.4, range 0-50). The mean number of years since onset of manic episodes was 15.1 (s.d. 11.3, range 1-42), and mean years since depressive episodes was 19.0 (s.d. 12.9, range 1-43). Given the large range in the data, median values may provide a more representative summary of the data. The median number of episodes was three for both depressive and manic episodes, while years since depressive onset was 17 and years since manic onset was 12.

In order to formally test whether self-selected samples differ from solely psychiatrically recruited samples, it would be necessary to carry out a systematic meta-analysis, which is beyond the scope of the current thesis. However, in order to informally assess whether the current sample was unusual, I decided to compare it to other samples from the DTI literature in bipolar disorder (excluding first episode and paediatric samples), as these studies are of particular interest. Table 2.1 shows the average number of episodes, age and duration of illness since onset for these studies. From this brief survey, it is clear that, at least within the DTI bipolar literature, most (7/10) studies have taken the pragmatic route of recruiting both from the community and psychiatric services – with only 3 recruiting solely via psychiatric contact. Unfortunately, only three studies provided data on number of affective episodes[10], and this data was limited in nature (one study providing only the range of episodes, one study reporting hospitalisations and one reporting only a percentage of subjects with more than 3 manic episodes). It is perhaps interesting that of the 7 studies that provided information on duration since onset, the three psychiatric-contact-only studies had the longest duration since onset. However, as this is not a systematic review, and because both duration of onset data and number of episode data is confounded by age, it is not possible to conduct a statistical analysis or draw firm conclusions.

[10] Possibly due to the inherent difficulty in establishing accurate numbers of affective episodes in patients with bipolar disorder. For patients with only a few, well defined episodes this is not problematic, however, for patients with numerous, or poorly defined episodes, it is difficult to accurately assess the number of prior episodes.

Table 2.1 Clincial Characteristics of Patients From DTI Studies of Bipolar Disorder

Study	Manic Episodes	Depressive Episodes	Age	Duration Since Affective Onset	Recruitment Method
Wang et al[140] 2008	-	-	32.6	-	Various.
Vercace et al[144]	-	-	35.9	12.1	No info
McIntosh et al[143]	-	-	39.3	18.8	Psychiatric caseload
Bruno et al[138]	No of episodes Range 1-35		39	13.8	Outpatient clinics and adverts
Houenou et al[137]	-	-	41.8	23.7	Psychiatry department
Yurgelun-Todd et al[133]	-	-	32.9	12.0	Outpatient
Regenold et al[132]	15 Hospitalisations		32	20.1	Inpatient sample
Beyer et al[131]	85% > 3 1-100	-	44	13.6	Community, not clear.
Haznedar et al[130]	-	-	43.9	-	Outpatient vlinics and previous study
Adler et al 2004[129]	-	-	32	-	Adverts

2.4. Inclusion and Exclusion Criteria:

The following inclusion and exclusion criteria were applied:

Inclusion:

- **All subjects:** Being a member of a twin pair, where both members are willing to take part in the study.

- **Controls**: No significant personal or familial (up to second degree relatives) psychiatric history. Significant history was defined as an evidence of psychotic, bipolar, schizophrenia or recurrent depressive illness.

  Note: due to its high prevalence in the general population, a lifetime diagnosis of Major Depressive Disorder – single episode was not considered to a reason for exclusion. Controls were, however, excluded if they were in a current depressive episode or had a diagnosis of Major Depressive Disorder, recurrent type.

Exclusion (all subjects):

- Age less than 16 or greater than 65.
- Organic brain disease
- Head injury resulting in loss of consciousness of greater than 10 minutes.
- Current (within 12 months) drug or alcohol dependence (according to DSM-IV criteria)
- Standard safety criteria for MRI

2.5. Subject Consent

Once potential participants volunteered to take part in the study, they were provided with complete information about what was involved and given the opportunity to ask questions. All participants were asked to provide written consent for their participation in the study and specific consent for access to medical records, blood samples, cheek swabs, as well as for participation in MRI scans. All participants were informed of their right to withdraw from the study at any time.

2.6. Assessment of Subject Characteristics

Initial Screening

Volunteers were given a brief telephone interview based on a screening questionnaire (include in appendix). This enabled an initial assessment of the volunteer's suitability for the study (questions included information about personal and family psychiatric history and major illnesses), and collection of socio-demographic information (gender, date of birth, ethnicity, years of education, volunteer's occupation and parental occupation).

When the initial screening indicated that the volunteer was not suitable for participation according to the inclusion and exclusion criteria of the study, the reasons were explained and the volunteer was thanked for their time. Otherwise, the volunteer underwent a detailed clinical interview, as outlined in *Clinical and Diagnostic Characteristics.*

Zygosity

Zygosity was preliminarily ascertained on the basis of a twin questionnaire. The results of the questionnaire were confirmed by DNA analysis of blood or cheek swab samples. DNA analysis was based on a set of 18 highly polymorphic markers (consisting of between 5 and 15 alleles and a mix of di, tri and tetra-nucleotide microsatellites). The results from each twin pair were compared to look for matching genotypes/alleles and a statistic calculated to determine the probability of the pair being monozygotic or dizygotic.

Clinical and Diagnostic Characteristics

For all patients and controls, clinical assessment was performed using the Schedules of Clinical Assessment in Neuropsychiatry (SCAN 2.1). All assessments were performed by one of two researchers who were formally trained in SCAN assessments (Dr Sridevi Kalidindi and the author). In cases where the correct diagnosis was uncertain at the end of SCAN assessment, both the patient's relatives and the medical records were also consulted.

Mood symptoms at the time of neuropsychological assessment were assessed in all participants using two subjective mood scales – the Beck Depression Inventory (BDI)[164] and the Altman Self Rating Mania Scale (ASRM) [165] which rate mood in the

week prior to administration. In addition, on the day of the MRI scan, two objective mood rating scales, the Hamilton Depression Scale (HAM-D)[166] and the Young Mania Rating Scale (YMRS)[167] were administered.

The HAM-D and YMRS were the primary measures chosen to assess affective status on the day of scanning, but these scales require the presence of a trained observer. In the case of the MRI scans, such a researcher (either Fergus Kane or Dr Sridevi Kalidindi) was always present, but this was not always the case on the day of neuropsychological assessment. Thus the HAM-D and YMRS were supplemented with the BDI and ASRM. The BDI and ASRM are both relatively quick, self-report measures, while the HAM-D and YMRS are rated by a trained observer, by way of an interview with the participant. Thus, the HAM-D and YMRS scales arguably provide a more thorough and objective assessment of a participant's mood and were therefore used in preference to the BDI and ASRM for neuroimaging analysis.

Socio-Economic Status

Socio-economic status was based on parental occupation at the time of the subject's birth (as ascertained during the screening telephone interview). Classification was based on the widely used system utilised in the Standard Occupational Classification, 1991[168]. Where both parents were in employment, the highest social class was used.

Handedness

Handedness was assessed by the Annett Handedness Questionnaire[169]. Although, handedness represents a potential confound in brain imaging studies (eg Hater, 2007[170]), due to the rarity of the patient sample, both dextral and sinistral volunteers were accepted. Handedness was therefore treated as a covariate in relevant analyses.

Medication

Participants were asked to detail their current psychiatric and non-psychiatric medication at the date(s) of testing. All patients had been stabilised on their medication for at least one month prior to scanning. To the author's knowledge, there is no analogue of the chlorpromazine equivalency for mood stabilising drugs, so no attempt has been made to calculate such a value.

2.7. Neuropsychological Assessment

A comprehensive neuropsychological battery was conducted with each participant. The battery was administered by three trained neuropsychologists: the author, Dr Eugenia Kravariti and Ms Anna Georgiades.

The battery consisted of the following tests:

- Wechsler Abbreviated Scale of Intelligence (WASI), 4 subtest version.
- California Verbal Learning Test (CVLT)
- Stroop
- Emotional Stroop
- Cambridge Neuropsychological Test Automated Battery (CANTAB). Six subtests:
    - Motor Task
    - Paired Associates Learning
    - Pattern Recognition Memory
    - Spatial Working Memory
    - Intra Dimensional-Extra Dimensional Shift
    - Rapid Visual Processing
- Verbal Fluency (category – animals and phonological – letter 's' only).

2.8. MRI Data Acquisition

All MRI scanning was conducted in the Mapother House Scanner at the Institute of Psychiatry campus of King's College, London. For all participants, data were acquired using a 1.5 Tesla GE N/Vi Signa System scanner (General Electric, Milwaukee, WI, USA), with actively shielded magnetic field gradients (maximum amplitude 40 mT m-1). A standard quadrature birdcage head coil was used for both RF transmission and NMR signal reception.

Participants were scanned in two sessions to minimise fatigue and claustrophobia. The first session consisted of four functional acquisitions and one diffusion tensor acquisition (DTI). The order of the functional acquisitions was randomised, but the DTI was always last. The second session consisted solely of structural image acquisition.

The specific acquisition parameters are detailed in the relevant experimental chapters.

2.9. Data Analysis:

The specific data analysis and methods are discussed in the relevant chapters. There are however, some issues with are common to the statistical analysis adopted in each chapter and these are discussed below.

2.9.1. Addressing Observation Dependence in the Statistical Analysis of Twin Data

Standard statistical methods of assessing differences between groups, such as analysis of variance (ANOVA), assume that the sampled data is independent. However, twins and family members share both genetic and environmental non-random factors and thus are not independent. In the standard ANOVA treatment, each observation is assumed to be independent from other observations. This is an assumption that does not hold true for all study designs. Repeated measures and family designs both violate this assumption; in a repeated measures design, the multiple observations for the same subject are likely to be highly correlated, while in a family design, the shared genetics and environment are also likely to result in correlation between the scores of the members of a single family. According to Myers and Well[120] is it is likely that violation of the independence of observation assumption, in the absence of specific measures to model dependence of observation, will increase the probability of type I (false positive) errors.

One solution is to employ cluster based statistics[11], as implemented in the statistical package STATA, which treats each twin pair or family unit as a cluster (thus each family member is treated as non-independent, but each pair/family cluster as independent). The cluster method used by STATA calculates a robust estimate of variance known as the Huber-White-sandwich estimate of variance[171,172]. For a more detailed explanation of the use of the Huber-White-sandwich estimate in cluster statistics, please see Williams (2000)[173].

While the above approach has been used successfully in the statistical analysis of neuropsychological data[174], there are currently no brain-imaging packages that allow for appropriate treatment of family or twin data within a whole brain analysis approach. The latter requires the use of specific analysis techniques built into brain-imaging packages such as SPM and XBAM. Theoretically, it is possible to adapt such packages to account for non-independence, but this would be a major undertaking and thus falls out of the scope of this thesis (in contrast, it would be logistically easier and feasible to extract region of interest data from the brain images and analyse with cluster analysis).

A potential solution is to use ANOVA within brain imaging software to identify group differences, extract the statistical values of these areas for each subject and then rerun the ANOVA in STATA using cluster statistics in order to account for non-independence. This, however, would be statistically problematic. The primary reason for this is that it would not be possible to account for the corrections for multiple comparisons that are applied by packages such as SPM and XBAM and thus a suitable statistical threshold could not be selected for subsequent statistical tests. Specifically, because areas would have already been selected after extensive

---

[11] It should be noted that the term 'cluster analysis' is also commonly used, in a different sense, in analysis of neuroimaging data (including the analyses described in this thesis), where the 'cluster' refers to spatially contiguous regions within an image. While potentially confusing, in practice the distinction between these two usages should be clear from context.

correction for multiple testing, the statistical significance of the extracted data would be distorted. An approach that might be useful would be to extract the data fro a specific brain area for analysis in STATA and then run to analyses both accounting for, and not accounting for, the familial clustering in the data. This would enable one to draw conclusions about how familial clustering affected the results in this area. However, one could not necessarily generalise from this area to other areas.

Another option would be to modelling twins as a repeated measure within a family. This is a reasonable solution to the problem of non-independence. However, while this is useful for twin-cotwin analyses, it is not easy to implement with more complicated data structures, especially in software packages such as SPM. In the full sample from the present study, not only are subjects non-independent between groups (i.e. affected twins vs. unaffected twins), but they are also non-independent within groups (concordant bipolar pairs in the patient group and control twins pairs in the control group). As far as I am aware, neither SPM, not other comparable software packages such a XBAM/CAMBA provide the flexibility for simultaneously modelling both within and between group non-independence of observation. In future this may not be the case, as at least one group (at University College, London) has declared an intention to address the issues of modelling twin and family data within SPM (personal communication).

Confronted with this methodological problem, I decided to adopt a two stage solution to the problem of non independent subjects in whole brain imaging. Essentially, this involves testing two separate general hypotheses:

1. Individuals with BD will show localised differences (of BOLD or DTI parameters) relative to controls

2. Relative to controls, non-bipolar co-twins of twins with bipolar disorder will show differences in the same areas as are identified in hypothesis one.

Step one can be tested using a whole brain ANOVA within a standard brain-imaging package using INDEPENDENT subjects (i.e. only one twin from each twin pair concordant for BP and one twin from each control pair). For step two, using ALL subjects, statistical data can be extracted from the areas identified in step one and a robust ANOVA can be performed in STATA. Step one is valid statistically as the

samples are independent. Step two tests a separate, independent hypothesis and thus is also statistically valid.

The method outlined here is not ideal, as it does not allow the whole sample to be used in the initial analysis. However, given the limited options currently available, I considered this to be the most appropriate method.

Following the steps outlined so far, if differences are found between the non-bipolar co-twins of twins with bipolar disorder and controls, a final analysis can be carried out in order to ascertain whether these differences are larger between DZ than between MZ twin pairs. This final analysis allows one to infer to what extent such differences are influenced by genetic makeup, as opposed to environmental influence.

Alternatives To STATA's Robust Regression with Clustering.

Multi-level modelling was considered as an alternative to using STATA's robust ANOVA with familial clustering. However, the advice from the Maudsley Bipolar Twin Study statistician was that, given the limited sample size, multi level modelling was unlikely to prove advantageous. However, for future analyses, stemming from the current thesis, in which both twin and family study data will be included, providing a larger sample size, it is likely that a multi-level modelling approach will be adopted.

2.9.2. Random Selection of Twins From Twin Pairs

Random selection was carried out in STATA. To do this, each twin from a twin pair was manually assigned as twin 1 or twin 2 (according to the order in which they had been added to the subject database). For each twin pair a random number between 0 and 1 was chosen using the stata 'runiform()' command. Where the number was < 0.5 twin 1 was chosen, where the number was >=0.5, twin 2 was chosen.

Data Wastage and Instability of Results.

By selecting only one member from each twin pair, the potential power of the study is inevitably reduced. However, given that no other suitable analysis method was identified, this was considered to be the best available option in order to conduct a statistically rigorous analysis. It is also true that an alternative selection of subjects from the available data pool might have resulted in different results. However, this criticism applies to any study, in that alternative selection of participants from the

wider data pool (i.e. the general population) could result in significant changes to the results.

2.9.3. Issues of Twin Designs and Birth Complications.

An unavoidable, but significant complication in the use of twins in research is that they are undeniably subject to a more complicated birth environment than singletons. The issues range from a higher chance of obstetric difficulties through to complications such as twin-twin transfusion than are exclusive to multiple births. For instance, while accounting for less than 2% of all births, multiple pregnancies (of which circa 98% are twins) account for between 9 and 12% of all perinatal deaths[175]. Multiple pregnancies are associated with higher rates of almost every potential complication and antepartum complications are seen in 80% of multiple pregnancies, as compared to 25% of singleton pregnancies[175]. It is likely that such increased risks of obstetric complications may lead to developmental differences in twins compared to singletons. Evidence for such differences is limited, but a recent meta-analysis by Voracek et al[176], found evidence that being a twin is associated with reduced IQ. A weighted average of this difference over all studies gave a best estimate of 4.2 IQ points. This difference is less than 1/3rd of standard deviation, but it is still significant. Interestingly this difference is greater in older studies than newer studies, which might suggest that improvements in anti-natal care have helped ameliorate the negative effects of obstetric complications.

Given the above, it is clear that caution needs to be exercised when generalising from studies of twins to studies of the general population. However, twin status may be treated in much the same way as any other demographic variable and potential confounds reduced by matching groups. Thus by using a twin sample for both the case and the control groups, it is possible to minimise the potential confound of twin status. In such a study, one may reasonably conclude that any statistically significant differences between groups are genuine, rather than simply due to twin status. Nevertheless, the possibility remains that twin status may still affect the results of the study via interaction with, for instance, diagnostic status. Thus replication of interesting results in a singleton sample (where possible) remains important.

2.10. Descriptive Characteristics of the Extended Participant Sample.

The socio-demographic and mood characteristics of the extended twin sample (all twins who participated in at least one of the studies outlined in chapters 3-5) are summarised in Table 2.2 (the specific characteristics of the sub-sample included in the statistical analysis of each chapter are described separately in the relevant chapter).

Following the exclusion of three participants (see below), the extended study sample included 112 participants, consisting of 54 twin pairs and 4 twins without their co-twin, falling in one of the following categories:

- 15 MZ discordant pairs (with one missing BD twin)
- 10 MZ concordant pairs (with two missing twins from two separate pairs)
- 18 MZ control pairs
- 8 DZ discordant pairs
- 1 DZ concordant pair
- 7 DZ control pairs (with one missing twin)

Four participants completed other parts of the study protocol, but did not provide neuroimaging data: two (one BD twin from a discordant pair and one BD twin from a concordant pair) were unable to enter the scanner due to claustrophobia, one BD twin from a concordant pair could not be scanned due to Tardive Dyskinesia, and one control twin refused to come to the Institute of Psychiatry for a scan.

2.10.1. Excluded Subjects

Three subjects were initially entered in the study but were later excluded. Two were members of the same pair, and the established diagnosis for the affected twin (based on the SCAN interview in the study) was autistic disorder with psychosis rather that bipolar disorder, as the family had informed the researchers. The third subject received a diagnosis of BD-II both from her psychiatrist and as part of the SCAN interview, but a later inspection of her medical notes revealed a differential diagnosis of Klein-Levin syndrome.

2.10.2. Low IQ of Concordant Twins Relative to Discordant Twins

In this study, the average IQ of twin pairs concordant for bipolar disorder is considerably less than that of control subjects and twins from discordant pairs (see table 2.1). Comparing across MZ concordant bipolar and MZ discordant bipolar groups (there was only one concordant DZ pair, so these were excluded from the analysis), this difference in IQ was significant ($p=0.017$). There was no significant difference in gender, socio-economic status, age, years of education, onset of bipolar disorder, years since onset, episodes of mania or depression or current levels of mania or depression. There was however a significant difference for gender (Pearson $chi^2=0.03$), with 75% of concordant MZ bipolar patients being male, compared to 21% of discordant MZ bipolar patients. There was no significant relationship between gender and IQ.

The finding of lower IQ in the concordant pairs is interesting and it is unfortunate that previous twin studies in bipolar disorder do not provide comparable data. It is possible that twins from concordant pairs have a higher genetic loading towards bipolar disorder and/or have experienced more adverse environmental effects than those from discordant pairs. This is not reflected in the clinical and demographic characteristics of the sample, as apart from in gender, there are no significant between group differences. It is possible that an increased genetic or environmental burden may affect IQ more than it does other clinical or demographic characteristics. However, in the case of bipolar disorder, there is no good evidence that the disorder is linked with significantly reduced general IQ (see section 1.3). Neither is there any convincing evidence that IQ and bipolar disorder share genetic pathways. Thus, it does not necessarily follow that higher genetic or environmental loading to bipolar disorder would be associated with reduced IQ. The between-groups difference in gender is dramatic, but it is unlikely that this is responsible for the difference in IQ. One previous study[177] has found an interaction between gender and cognitive performance in bipolar disorder, but in this case females had greater deficits than males in the domains of verbal learning and memory and no deficits in IQ were reported. There are no other reports of gender X cognitive function interactions in the bipolar literature. It should also be acknowledged that the sample size in the concordant group is rather small, being just 8 pairs of twins. Nevertheless, as this

finding is potentially very interesting, it would be useful to investigate whether it is replicated in other, larger samples.

2.11. Problems with ANCOVA when groups differ on the covariate.

The following is a explanatory note on issues associated with covariation in psychiatric studies. This will be referred in future chapters.

Including covariates in analyses is a commonly used way of improving the power of a test of the independent variable by accounting for extra sources of variance. When the covariate shares no variance with the grouping variable, entering the covariate into a regression along with the group results in variance being associated with the covariate, thus reducing the variance associated with the grouping variable, allowing more subtle effects to be detected. This is the ideal situation and occurs in the case of true random selection (as might be seen, for example, in a placebo/treatment trial of control subjects).

In such situations, the use of ANCOVA is encouraged and may be a very useful tool. However, according to Miller and Chapman, analysis of covariance is commonly misinterpreted. In their 2001 paper, 'Misunderstanding Analysis of Covariance', the authors argue vociferously (and are backed up by a formidable literature) that it is not appropriate to conduct ANCOVA when groups differ on the covariate. Indeed, the authors go further and (quoting previous authors) and argue that 'there is no statistical method that can address the question of whether two groups that differ on variable A would differ on variable B if they did not differ on variable A'. There are however, situations in which this rule may be too strict. If two non-randomly selected groups differ on variable B due entirely to chance (and thus the grouping variable and the covariate do not share variance), the situation is essentially as with truly random selection, and covariation is justifiable.

Various mathematical treatments of this issue (known as the 'Lord's Paradox') are available, but Miller and Chapman attempt to explain the issue in a more theoretical manner and by providing examples. The essential argument is that following removal of a covariate that is related to the grouping variable, it is not clear exactly what the

residual of the grouping variable represents and thus group is no longer a good measure of the construct that it is supposed to represent. Regression adjustment may remove part of the group effect or equally may introduce a spurious group effect. The authors site a specific example very relevant to the current thesis and its associated literature:

'If we compare a sample of depressed patients with non-patient controls and covary out anxiety, which happens to be higher in the patients, it is not necessarily the case that the residual group difference is a clear, clean representation of depression as it would exist without the comorbid anxiety. What we should believe about that depends on our model of the relationship between depression and anxiety. If they happen to co-occur because of non-specific severity factors that are not specifically related to depression, our ANCOVA might be effective in removing such variance, leaving 'pure' depression. If however, we believe that the negative effect that depression and anxiety share is central to the concept of depression, then removing negative affect (by removing anxiety) will mean that the group variance that remains has very poor construct validity for depression.'

The authors provide a variety of other examples to illustrate their arguments and the paper provides an excellent and readable overview of the issues.

Moving to the example in hand, it is tempting to attempt to covary for subsyndromal depressive and manic symptoms in the current sample. We would do this in order to observe the relationship between bipolar disorder and brain activation, without the confounds of affective symptoms. However, we face the same problem detailed above, as our groups differ on their affective scores precisely because of the way we have selected them; their group membership is thus intimately related to their affective scores. Statistical methods cannot remove the effect of affective symptoms from bipolar disorder because they are so intimately linked.

It should be noted that Miller and Chapman do discuss a variety of strategies for addressing this fundamental issue in psychopathology. However, these strategies are generally not applicable to the current study. Further, while the above issues are important for between group analyses, they do not preclude investigation of the covariate as a substantive variable in within group analyses.

Table 2.2 Demographic characteristics of the total sample

	Monozygotic				Dizygotic			
	Discordant		Concordant	Control	Discordant		Concordant	Control
	BD	Unaffected	BD		BD	Unaffected	BD	
Demographics								
N	14	15	16	36	8	8	2	13
BD I, BD II	13,1	NA	16	NA	7,1	NA	2	NA
Gender (% male)	21	20	75	22	38	38	0	23
Ethnicity (% white cauc, other)	93,7	93,7	88,12	94,6	100,0	100,0	100,0	69,31
Handedness (% left, right, mixed)	79,14,7	93,7,0	81,13,6	89,6,6	88,13	88,13	100,0,0	85,15,0
PSC (% I, II, III, IV, V, UE)	21,21,43,7,7	20,27,47,7	0,13,37,25,25	17,33,31,6,8,6	25,63,13,0,0	25,63,13,0,0	0,0,100,0,0	15,39,39,8,0
Age (years) mean, (sd, range)	40.4 (14.5,21-61)	41.2 (14.2,21-61)	41.4 (13.6,22-63)	37.1 (11.0,22-56)	39.5 (12.0,27-55)	39 (12.0,27-55)	56, (0,56-56)	31.5 (11.4,20-51)
IQ mean (sd, range)	113.9 (10.1,99-128)	114.3 (8.9,95-127)	102.2 (12.7,84-138)	113.5 (12.1,81-129)	119.8 (8.6,109-134)	123.5 (6.8,113-136)	109.0 (5.66,105-113)	120.0 (7.0,107-129)
Years of education mean (sd, range)	15.6 (2.5,10-19)	16.3 (2.8,10-22)	13.9 3.2,10-22	15.2 (2.6,10-19)	17.0 (2.6,13-21)	16.9 (2.0,13-19)	12.0 (0,12-12)	16.4 (2.1,12-29)
Onset of BD I / BD II mean (sd, range)	22.9 (6.9,14-37)	NA	25 (10.6,14-55)	NA	22.8 (7.1,13-34)	NA	29.5 (4.9,26-33)	NA
Years since onset of BD mean (sd, range)	18.5 (12.9,3-40)	NA	17.7 (12.8,3-40)	NA	17.1 (11.7,5-42)	NA	26.5 (4.9,23-30)	NA
Episodes of Mania/Hypomania mean**,** median (sd, range)	2.5, 2 (11.7,0-40)	NA	12.9, 5 (25.5,2-100)	NA	9.6,4 (10.5,1-31)	NA	3,3 (1.4,2-4)	NA
Episodes of Depression mean**,** median (sd, range)	7.4, 2 (11.7,0-40)	.5, 0 (.8,0-3)	5.7, 3 (5.9,0-19)	0.0, 0 (.2,0-1)	10.6, 2 (18.1,0-50)	0.0 (0,0-0)	15, 15 (14.1,5-25)	.2, 0 (.4,0-1)
Mood								
HAM-D (sd, range)	8.6 (8.7,0-27)	2.3 (2.4,0-7)	5.1 (6.2,0-17)	0.6 (1.4,0-5)	3.9 (8.0,0-22)	0.9 (1.5,0-4)	2.5 (3.5,0-5)	1.1 (1.7,0-7)
YOUNG-M (sd, range)	3.2 (4.0,0-12)	1.5 (3.0,0-10)	1.1 (2.3,0-9)	0.5 (1.0,0-3)	1.0 (1.9,0-5)	0.3 (0.8,0-2)	0.5 (0.7,0-1)	0.4 (0.9,0-3)

Key to Abbreviations: PSC=parental social class (groups I-V, unemployed). HAM-D: Hamilton Depression Scale. Young-M: Young Mania Scale

3. A fMRI Investigation of Brain Activation in Twins with Bipolar Disorder during a Verbal Working Memory Task.

3.1. Introduction

In this chapter I present an analysis of the N-Back verbal working memory fMRI data from the Maudsley Bipolar Twin Study. Much of the literature that forms the rationale for this study has already been discussed in the introduction to the thesis, so what follows is primarily a summary. The reader is encouraged to refer to the relevant section of chapter 1 for a more detailed overview of the literature.

It is now reasonably well established that bipolar disorder is associated with cognitive impairment, and that this impairment is present not only in the manic and depressive phases, but also during symptom remission. Such cognitive impairment is reported across in a variety of domains; the strongest and most consistently reported deficits are reported in the domains of executive function, memory and attention[178,25,26]. Deficits in verbal working memory and executive function have the largest effect sizes, which, although they vary depending on the specific variable, are mainly in the 0.5 – 1.0 range. Not only are these deficits present in euthymic patients; they are also present (albeit in attenuated form) in the unaffected relatives of bipolar patients[178], which marks them as potential endophenotypes for bipolar disorder.

Having identified such deficits, it is logical to try and identify the biological mechanisms that may underlie them. Functional Magnetic Resonance Imaging (fMRI) provides a powerful way to investigate the neural correlates of such deficits. Two tasks have been used to investigate working memory in bipolar disorder, the N-Back and the Steinberg task; the results from both tasks have been equivocal.

Four fMRI studies of bipolar disorder have been published using the N-Back paradigm[115,114,119,116]; of these, one study[116] found no significant differences between patients and controls, while three found significant group differences [115,114,119]. Among the latter, the first[115] found increased bilateral frontal, temporal and parietal activation, with decreased left precentral, right medial frontal and left supramarginal activation. The second found mainly increased activation (bilateral prefrontal, temporal, posterior parietal gyri and thalamus), with decreased activation only in the posterior cingulate. The last of these studies, which investigated both patients with bipolar disorder and their unaffected relatives, found significantly increased activation in the left frontal pole of unaffected relatives, and a trend towards increased activation in the same area for patients. Hyperactivation of this area was also seen in an earlier

study by Adler et al[114] and represents the only replicated finding in the published studies to date. Importantly, the finding of altered activation in unaffected relatives also suggests that prefrontal hyperactivation could represent an endophenotype for bipolar disorder. However, as discussed in chapter 1, twin studies are required to demonstrate that such differences represent true endophenotypes.

There are also three papers published using the Sternberg paradigm[xii], one of which found no activation differences between groups[115], and two of which (using the same sample, but different analyses) found widespread decreases in activation relative to controls[121,125]. In the later two studies, the task involved emotional verbal stimuli, which means that it is difficult to say whether activation differences were due to the working memory task, emotional valence or an interaction of the two.

It is not clear exactly why these studies have found such differing results, but small sample sizes, divergent methodologies, medication effects and differences in the patient characteristics may all play a part.

The current study aims to add to this literature, and extend it by using twin methodology to identify whether differential activation during a working memory task may represent an endophenotype for bipolar disorder. Investigating both patients with bipolar disorder and their unaffected co-twins can also address the potentially confounding effects of medication and residual sub-syndromal mood symptoms on differences between patients and volunteers.

[xii] In the Sternberg task, volunteers are first shown a list of stimuli. They are then presented with probe items and must identity whether on not the probe items were in the list.

3.1.1. Hypotheses:

Based on the literature extant at the start of the study the following hypotheses were tested:

1. During a verbal working memory task, twins with bipolar disorder will show altered activation relative to control twins in brain areas normally engaged by the paradigm. These differences will be more evident as memory load increases.
2. Qualitatively similar, but quantitatively less severe differences in brain activation observed in bipolar patients will also be seen in their unaffected co-twins.

3. These intra-twin pair differences will be more pronounced in fraternal than in identical twins, reflecting the genetic basis of these differences.

3.2. Materials and Methods

3.2.1. Participants

Participants were drawn from the sample described in chapter 2 (methods). The specific characteristics of the sub-samples used are described in the relevant results sections.

3.2.2. Data Acquisition

Single-shot, gradient echo imaging was used to acquire 270 T2*-weighted image volumes on a neuro-optimized 1.5 Tesla GE LX System (General Electric, Milwaukee, Wisconsin) at the Maudsley Hospital, South London and Maudsley NHS Trust, London, United Kingdom. For each volume, 16 non-contiguous axial planes parallel to the intercommissural plane were collected with the following parameters: repetition time (TR) = 2000 msec, echo time (TE) = 40 msec, slice thickness = 7 mm, slice skip = .7 mm, in-plane resolution = 3 × 3 mm.

At the time that the study was started, and with the scanner available, the protocol used for the N-Back MRI data acquisition represented the best compromise between slice thickness, brain coverage and repetition time (there is always a trade off between these parameters, and the protocol used depends on which are considered most important). Interslice gaps are necessary in order to prevent 'cross-talk', a phenomenon whereby, due to imperfections in RF pulses, acquisition of signal from a slice interferes with that of neighbouring slices[179]. In this study, cross-talk is minimised by a combination of interleaved acquisition (acquiring first even, then odd numbered slices) and interslice gaps.

While, the protocol was arguably optimal at the time the study was started, both the slice thickness and the thickness of the interslice gaps may be considered as limitations to the current study. Thus, if the study was to be rerun today and a 'state of the art' scanner was available, I would aim to use a protocol specifying both thinner slices and smaller gaps.

3.2.3. N-Back Verbal Working Memory Task

Rationale For Task Choice

At the time of the inception of the current study, deficits of verbal working memory and executive function were among the most promising candidates for endophenotypes of bipolar disorder. Based on this, it was considered that verbal working memory and executive function represented ideal targets for an fMRI investigation of the neural correlates of bipolar disorder.

Rather than design a new fMRI paradigm from scratch, it was decided that a well established paradigm should be used. The two primary candidates for a verbal working memory paradigm were the Sternberg and the N-Back tasks. The Sternberg task[180] is a test of verbal memory storage and recognition capacity, processes that reply on the phonological loop[181]. In the task, participants are presented with a list of words; they are then shown a probe word and asked to indicate whether they have seen it before. The N-Back task[182] is a continuous performance task that assesses working memory and manipulation of stored information, processes that require executive control. The task requires subjects to store between one and three letters (the 'N' in N-Back refers to the variable number of letters to be stored) in memory while continually updating the contents of their memory, (adding new letters and dropping the old letters) and comparing the current stimulus with previous stimuli.

A possible advantage of the Sternberg paradigm over the N-Back is that it may engage more 'pure' cognitive processes than the N-Back as it does not involve a continuous performance aspect, and involves separate encoding and recognition phases. However, the Sternberg does not involve significant engagement of the central executive working memory component, instead placing greater demands on the phonological loop. The N-Back, by contrast, involves simultaneous encoding, manipulation and recognition processes, placing heavy demands on the central executive working memory component. Nevertheless, as pointed out by Monks et al[115], 'it should be acknowledged that no task unequivocally reflects one cognitive process or operation'; the Sternberg task is not free of demands on executive resources, while the N-Back also recruits the phonological loop to some extent.

Although the N-Back is a continuous attention task, by parametrically varying the working memory component of the N-Back paradigm, it is possible to partially

disentangle the working memory and continuous attention components of the task. For instance, if task performance differs between groups at 1-back level and this difference does not increase at 2 and 3-back levels, one may reasonably conclude that the difference is primarily due to group differences in continuous performance ability. However, if differences are not detected at 1-back level, but emerge at 2 and 3-back levels, one may conclude that differences are primarily due to differential working memory abilities. Of course, it is also important to be aware of possible interactions between working memory and continuous performance abilities. For instance, participants, especially those who struggle with the task, may find it more difficult to maintain attention as cognitive load increases.

In case-control studies, where a high degree of participant heterogeneity is for practical reasons unavoidable, a reliable and robust paradigm is very desirable in order to keep extra variance to a minimum. The robustness and reliability of BOLD activation during the fMRI task was thus an important factor in deciding which task to use. At the inception of the study there had been significantly more published fMRI studies using the N-Back task than the Sternberg task (over 20 vs. 4 as of February 2003). Furthermore, the published N-Back fMRI results indicated that the paradigm consistently and reliably activates a brain network including bilateral frontal, parietal and temporal regions (as discussed more recently by Owen et al[183]).

The decision to use the N-Back task rather than other tasks such as the Sternberg task was also a pragmatic one. Ideally perhaps, both the Sternberg and the N-Back task would have been used (as in Monks et al[115]). This would enable comparison of two memory tasks, one with a significant central executive component, and the other without. However, when choosing from among the many potentially interesting tasks to include in the overall battery, it was important to take into account other factors including cost, scan time and subject fatigue. Thus only one fMRI memory task was included in the battery. The fact that the N-Back paradigm, but not the Sternberg, had previously been used successfully on the MRI scanner at the Institute of Psychiatry[184,185] was also a contributing (but not deciding) factor in the decision to use the N-Back. Overall, it was considered that the benefits of using the familiar and robust N-Back paradigm outweighed the potential benefits of either choosing an alternative existing paradigm such as the Sternberg or developing a new paradigm.

Task Presentation

During the task, participants lay on their back in the MRI Scanner. The task was presented to the participant via a projector screen, which they were able to see via a mirror mounted in the head coil. Participants responded by means of a MR compatible button box. The task was run on an IBM PC, running software coded at the Institute of Psychiatry, London. Button box responses and reaction times were recorded by the same software.

In all conditions, subjects were presented with a series of stimuli presented one at a time in the centre of the screen, at a rate of one stimulus every 2 seconds. In the baseline condition, the subject was presented with the following series: 'A' 'B' 'C' 'D' 'E' 'F' 'G' 'H' 'I' 'J' 'K'. The stimuli 'X' was pseudo-randomly inserted into this sequence (3 times) and the participant simply had to press the button box whenever they saw the letter 'X'. In the active conditions (known as 1-back, 2-back and 3-back), the participant was presented with a pseudo-random series of letters and was asked to look out for target stimuli (and press the button box when the see the target stimuli). The stimuli were as follows:

- **1-Back:** Any letter that is the same was the one presented immediately before it.
- **2-Back**: Any letter that was the same as the one presented two before it.
- **3-Back:** Any letter that was the same as the one presented three before it.

The task used a block design. Each block consisted of one instruction presentation and 16 stimuli. There were 18 blocks, 9 baseline and 3 for each of the active conditions. The blocks are presented in the following order: baseline, 1-back, baseline, 2-back, baseline, 3-back, baseline, 2-back, baseline, 1-back, baseline, 3-back, baseline, 2-back, baseline, 3-back, baseline, 1-back. Examples of each block are shown in Figure 3.1.

All subjects were trained on the task (using a different stimuli set) beforehand to make sure they understood the task and were able to perform it. Specifically each subject was (i) trained until they could correctly respond at the 2-back level and (ii) given the opportunity to attempt the 3-back stage. Correct response at the 3-back level was not required, but if subjects failed to respond correctly, the researcher established that this was not due to task difficulty and not misunderstanding of the instructions.

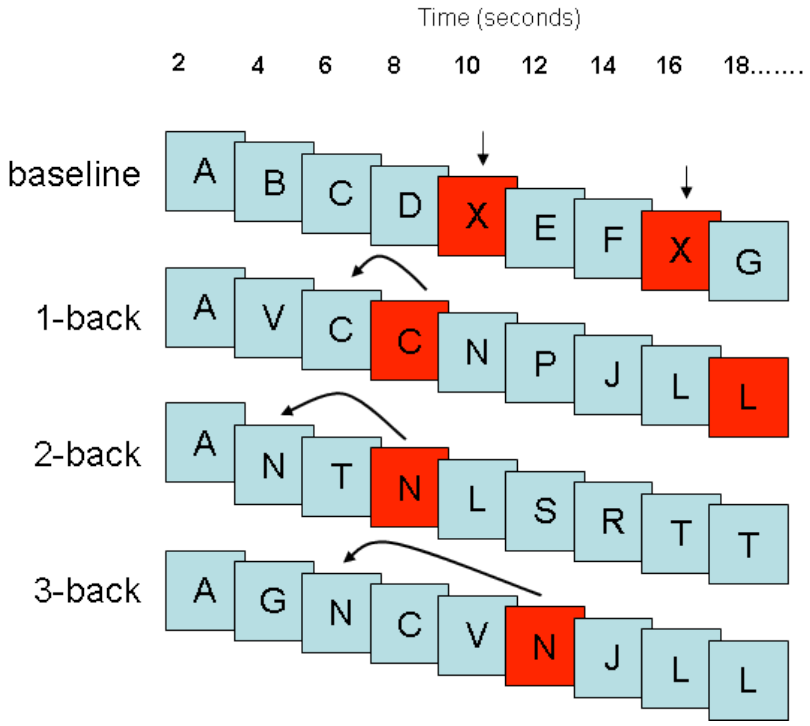

Figure 3.1 N-Back Stimuli Presentation

3.2.4. Data Processing and Analysis

MRI data were processed and analysed on the Institute of Psychiatry UNIX workstations using SPM 5 (Welcome Department of Cognitive Neurology, Institute of Neurology, London). Output from the MRI scanner was in UNC (University of North Carolina) format and thus data had to be converted to ANALYZE (Mayo foundation, Rochester, Minn.) format before further processing and analysis; this was done using the UNIX utility 'unc2spm'.

Non-independence of subjects. A multi-stage solution.

In order to avoid violating the assumption of independence of observations that is part of SPM's statistical analysis, it was necessary to utilise multi stage analysis solution (the rationale for this is discussed in more detail in chapter 2).

- **Stage 1. Preprocessing.** Scan data from all subjects were preprocessed to produce realigned normalised and smoothed images for each subject.

- **Stage 2. Analysis 1: Statistical Analysis of Group Maps in Independent Subjects**. In order to examine potential differences between twins with a diagnosis of bipolar disorder and control twins without a diagnosis of bipolar disorder, a group mapping technique was employed. Crucially, in this stage, only independent subjects were used.

- **Stage 3. Analysis 2: Region of Interest Analysis.** Stage 3 would have involved the analysis of extracted time series data from non-independent subjects. In the event, no group differences were detected in stage 2, therefore stage 3 was not conducted. However, so that the reader may fully appreciate the rationale, the proposed analysis is still detailed.

The three stages are described in more detail below:

Stage 1: Data Pre-Processing

As discussed in the methods chapter, when applying whole brain voxel based techniques, it is necessary to pre-process each subject's individual data before performing any group analyses. The following pre-processing steps were therefore performed prior to group analysis:

1. Realign and Unwarp: This step corrected for subject movement during the scanning process. This was run using standard SPM parameters.
2. Normalise: This step converted all subject scans into standard space. Here, the standard SPM EPI template was used as a reference image. All other SPM options were standard.
3. Smoothing: All images were smoothed to minimise the effects of inter-subject variability. Images were smoothed using an 8 x 8 x 8 full width at half maximum (FWHM) Gaussian kernel.

3.2.5. Advantages and Disadvantages of Event Related and Block Designs

The majority of fMRI studies can be categorised as either 'blocked' or 'event-related' designs. The primary difference between these two designs is the way in which stimuli are presented and analysed. Essentially in a block design, similar stimuli are presented within continuous epochs, or 'blocks'. A standard blocked design is the A-B design, in which two blocks (A and B) are alternated several times within a scanning session. Analysis of such a design may consist of simple t-tests to compare BOLD signal between A and B conditions, or may consist of more complex parametric or factorial tests.

Blocked design was inherited from PET methodology, in which event related designs were impossible due to the temporal resolution of PET. Functional MRI, by contrast, offers much better temporal resolution and thus allows more sophisticated designs. There are a number of potential problems with basic blocked designs. Such designs can be predictable and potentially tedious, leading to issues with habituation, strategy, anticipation and attention. It is also hard for a subject to maintain a continuous and pure cognitive approach during a task block. Equally, in rest blocks, a subject's brain will never be in a truly resting or neutral state. Partly as a result of these issues, hemodynamic response may vary within the block as a subject adapts or habituates to the task at hand. However, despite these problems, blocked designs remain useful due to the fact that they provide relatively robust results, good statistical power and relatively large BOLD signal changes[186].

In an event related design, rather than being grouped within blocks, stimuli are presented as unique events, which may be presented randomly (or pseudo-randomly) within the scan. In such a design, the hemodynamic response function can be estimated for the averaged data from multiple separate events. This opens the way to more sophisticated and less predicable experiments. Indeed, event related designs allow studies that would be near impossible in a block design (such as investigating the neural correlates of unexpected stimuli). Furthermore, with an event related design, individual specific differences in performance can be accounted for by post-hoc modelling of factors such as incorrect responses and response latency. Event

related designs have two main disadvantages compared to block design. The most serious is that single events result in significantly lower signal to noise ratios than do blocks of events. In order to compensate for this, it is necessary to increase the number of events within a scanning run. This in turn presents other problems, in particular maintaining subject interest (a problem in both design types). The second, less problematic disadvantage is that analysis of event related designs can be significantly more complicated and laborious than for blocked designs.

It is possible to apply event related analysis techniques to primarily block related designs[187]. This technique was applied in the current study. Within the N-Back task, participant's responses to individual stimuli (presented within blocks) were recorded. By modelling blocks as a collection of events, one can attempt to model variance in the BOLD signal due to variations in individual response. In the current study, responses (both correct and incorrect) were modelled as effects of no interest, removing (at least partially) the correlates of motor response, so that the remaining signal would be a 'cleaner' estimate of the neural correlates of cognitive monitoring of stimuli. This approach is necessarily a compromise, but enables the researcher to combine the more sophisticated modelling techniques from event related designs with the relative power advantages of blocked designs.

Stage 2: Whole Brain, Independent Subjects Analysis

Subject Selection

In order to avoid violating the ANOVA assumption of independence, subjects were selected such that no two subjects (within and across groups) were members of the same family. Thus in the BD group, all BD twins from discordant pairs were selected, while only one (randomly selected) twin was taken from each concordant pair. In the control group, one twin was selected from each pair; this selection was carried out randomly except when one of the pair met criteria for a psychiatric diagnosis. In such cases, the subjects who did not fulfil criteria for a psychiatric diagnosis were favoured[13].

Individual Subject Level Analysis (First Level)

Subject specific models were generated for each subject by convolving each onset time with a synthetic haemodynamic response function (HRF). The baseline, the 3 active conditions, correct and incorrect trials were all modelled separately using an event-related model.

In each condition there were relatively few response trials (8) relative to non-response trials (40), and these 8 responses were divided into correct and incorrect responses (false negatives); thus there was very little power to detect differences in the BOLD correlates responses/non response trials. Correct and incorrect responses were therefore modelled as effects of no interest (responses were still accounted for in the modelling of the variance, but could be removed from the later contrasts). There were four experimental conditions: (i) baseline monitoring (ii) 1-back monitoring (iii) 2-back monitoring (iv) 3-back monitoring. In order to reduce the possibility of user data entry errors, a MATLAB script (see appendix A1-5) was written to read individual subjects response files and specify the models automatically. In order to verify the automated technique, its output was compared against the results obtained

---

[13] According to the literature, there is considerable genetic overlap between psychiatric disorders such as bipolar disorder and depression. Therefore, given the relatively small sample size, it was considered optimal to have the most psychiatrically 'clean' control sample possible – and therefore the maximum psychiatrically mediated group differences. Such sampling does however, run the risk of selecting a non-representative ultra healthy control group.

from the equivalent manual procedure using 10 randomly chosen subjects. The results were found to be identical.

To remove low-frequency drift, the data were high-pass filtered using a set of discrete cosine functions with a cut off period of 128 seconds. SPM's implementation of the general linear model (GLM) was used to calculate parameter estimates for all brain voxels. Following model specification and estimation, for each subject, contrast images were generated for 1-back vs. baseline, 2-back vs. baseline and 3-back vs. baseline.

Group Level Analysis (Second Level)

The individual level contrast images were entered into a full-factorial ANOVA, permitting inferences at the population level. The ANOVA model was specified to include two factors: group (BD vs control) and task-load (1-back, 2-back, 3-back) and four covariates of no interest(age, IQ, gender and handedness). The group factor was specified as independent, while the task-load factor was specified as non-independent.

In this way, it was possible to generate second level contrasts in order to investigate the effect of group membership on activation in 1-back, 2-back and 3-back conditions (relative to the baseline condition), as well as the interaction between group and task difficulty. T images for each second level contrast were transformed into statistical parametric maps of the Z statistic. Difference regions were considered statistically significant if they survived family wise error[14] (FWE) correction at the level $p<0.05$.

---

[14] (P_{FWE} stands for probability, corrected for familywise error. A Familywise Error is (in the case of an SPM analysis) a false positive anywhere in the statistical parametric map volume created as the result of a contrast calculation. Thus, controlling for FWEs at $p<0.05$ means that for every 20 contrasts generated, one false positive would be expected. FWE correction at $p<0.05$ may be considered a conservative approach, however, given that an imaging study typically involves generation of multiple contrasts, each with a 1/20 chance of a false positive, the conservative approach is arguably appropriate. SPM also offers an alternative error correction method known as False Discovery Rate (FDR), which, rather than controlling for the chance of any false positives, controls the proportion of false positives expected among suprathreshold voxels. Thus, with an FDR of 0.05, for each contrast, 5% of discoveries are likely to be false. SPM's FDR approach is more sensitive than its traditional FWE approach, but this is at the expense of weaker control of familywise errors).

Masking

In order to maximise power to detect differences, second level contrasts were initially conducted using a mask – limiting the search space to brain areas activated during task performance. In limiting the search space in this manner, the level of necessary correction for multiple comparisons was reduced. The mask was generated from an 'average effect of condition' F contrast, which is automatically produced by SPM in full factorial analyses. Those regions showing a significant effect at $p<0.05$ (corrected for FWE) were entered in the full factorial analysis as an explicit mask. This mask therefore includes all areas demonstrating either significant increases or significant decreases in activation during task performance.

The mask chosen was generated based on the average effect of the N-Back task on BOLD response, so as to select only those areas clearly involved in the task response (both activation and deactivation). However while masking for the task related network increased the power to detect differences, it meant that any patient-control differences outside the masked region would be missed. In order to check for this, the analysis was also conducted without masking.

Stage 3: Analysis 2: Statistical Analysis of Group Maps in Independent Subjects.

As previously noted, in the event, no group differences were detected in analysis 1 and therefore no stage 3 analyses could be conducted. Had group differences been detected the following analyses would have been conducted: A new subsample would have been selected, consisting of all discordant twin pairs and all control pairs. This subsample would have been divided into three groups: BD twins from discordant twin pairs (DB), non-BD twins from discordant twin pairs (DNB) and control twins. For each subject, data would have been extracted for each region of difference identified in stage 2. Regression analysis (adjusted for familial clustering) would have then been conducted to investigate BD-NBD and NBD-Control differences in the identified regions. If differences were detected at this stage, a further analysis would have been conducted to see if differences between discordant bipolar twin pairs were greater in DZ than in MZ pairs.

Analysis of Demographic, Clinical and Behavioural Data

Statistical analysis of the behavioural data was conducted with STATA v10. Group differences on categorical data (such as ethnicity) were investigated with either Chi squared or Fisher's exact tests. Group differences on continuous data (such as age, IQ, response accuracy and response time) were investigated using regression analysis.

3.3. Results

3.3.1. Demographics, Mood and Clinical Characteristics

Table 3.1 Demographic, Mood and Clinical Characteristics.

	GROUP		Sig	Test	C.I.		Coef.
	BD	CONTROL			Low	High	
Demographics							
N	31 (29 BD I, 2 BD II)	24					
Gender (% male)	39	20	0.155	C2			
Ethnicity (% white cauc, other)	93.5, 6.5	88, 12	0.857	FE			
Handedness (% left, right, mixed)	81, 13, 6	88, 4, 8	0.563	FE			
PSC (% I, II, III, IV, V, UE)	13,29,35,13,10,0	13,33,33,4,8,4	0.858	FE			
Age (years) (sd, range)	41.8 (13.93, 21-63)	35.54 (11.51, 20-56)	**0.092**	reg	-13.09	1.01	-6.04
IQ (sd, range)	112.58 (12.60, 84-138)	114.29 (11.77, 81-129)	0.609	reg	-4.97	8.39	1.71
Years of education (sd, range)	15.41 (3.03, 10-22)	15.54 (2.68, 10-19)	0.877	reg	-1.45	1.70	0.12
Onset of mania/hypomania (sd, range)	26.7 (11.1, 13-55)						
Years since onset. (sd, range)	15.1 (11.2 1-42)						
Mood							
HAM-D (sd, range)	6.17 (7.72, 0-27)	0.792 (1.44, 0-5)	**0.001**	reg	-8.59	-2.16	-5.38
YOUNG-M (sd, range)	2.07 (3.34, 0-12)	0.33 (.70, 0-2)	**0.016**	reg	-3.13	-0.34	-1.73

Key to Abbreviations: C2: Chi Squared, FE: Fishers Exact, reg: regression, PSC=parental social class (groups I-V, unemployed). HAM-D: Hamilton Depression Scale. Young-M: Young Mania Scale

Demographic and mood data are summarised in table 3.1.

Demographics:

The bipolar disorder group did not significantly differ from the control group on gender, ethnicity, handedness, parental social class, age, IQ or years of education. There was however a trend (p=0.092) towards a difference in age, with the bipolar disorder group (mean age 41.8 years) being older than the control group (mean age 35.5 years).

Clinical and Medication Characteristics.

Comorbid diagnoses for each group are shown in Table 3.2. The BD group consisted of 29 patients with a diagnosis of BD-I and 2 patients with a diagnosis of BD-II. 8 of the BD patients also met criteria for current co-morbid diagnoses. The average age of onset of mania or hypomania was 26.7, with a relatively large range of 13-55 years. On

average, patients had onset of mania/hypomania 15.1 years prior to the scan. Four patients had current alcohol dependence, one of whom also met criteria for cannabis abuse. Two patients had panic disorder with agoraphobia. One patient had panic disorder without agoraphobia and one had generalised anxiety disorder. Lastly, one patient had obsessive compulsive disorder. In the control group, one subject met criteria for panic disorder.

Of the 31 participants in the BD group, 21 were taking mood stabilisers, of which 4 were on lithium monotherapy, 4 were on non-lithium mood stabilising monotherapy, 7 were also taking antipsychotics and 6 were also taking antidepressants. Of the remaining patients, 6 were taking no medication, 2 patients were taking only antidepressants, one was taking both an antipsychotic and an antidepressant and one was taking an antidepressant and a benzodiazepine. No controls were taking psychoactive medications.

Table 3.2 Frequency of Current Comorbid Conditions in Bipolar and Control Samples.

Comorbid Conditions (BPAD) or Primary Diagnosis + Comorbid Conditions (Control)	Frequency	
	BPAD	Control
Alcohol Dependence + Cannabis Abuse	1	0
Alcohol Dependence	3	0
GAD	1	0
OCD	1	0
Panic Disorder w Agoraphobia	2	1
Panic w/o agoraphobia + GAD	1	0

Mood:

As expected, the two groups differed significantly on both the Hamilton depression and the Young mania scales, with the bipolar disorder group having higher depression and mania scores (6.2 vs 0.7 and 2.1 vs 0.3 respectively).

3.3.2. Behavioural Data

Table 3.3 Behavioural Data

	Group		Sig	C.I.		Coef
	BD	Control		Low	High	
Percent Correct Responses (s.d, range)						
AllConditions	82.7 (19.4,12.5-97.9)	93.0 (6.4,77.0-100)	0.008	0.66	4.25	2.46
Baseline	93.4 (17.0,12.5-100)	99.8 (0.9,95.8-100)	0.041	0.03	1.51	0.77
1-Back	88.3 (22.4,25.0-100)	95.8 (13.1,37.5-100)	0.125	-0.09	0.69	0.30
2-Back	76.7 (26.67,12.5-100)	89.6 (14.6,50.0-100)	0.008	0.19	1.23	0.71
3-Back	59.9 (25.1,12.5-87.5)	76.1 (19.9,37.5-100)	0.023	0.10	1.25	0.67
No. of False Positives	3.0 (3.2,1-16)	4.5 (10.6,1-48)	0.483	-1.41	2.94	0.76
Reaction Time (seconds (s.d))						
Correct Responses	.67 (.13)	.59 (.08)	0.003	-0.07	-0.02	0.04
False Positive	.79 (.36)	.90 (.33)	0.293	-0.05	0.16	0.05
Correct at baseline	.60 (.14)	.53 (.08)	0.020	-0.07	-0.01	-0.04
Correct at 1-back	.67 (.15)	.55 (.09)	0.001	-0.09	-0.03	-0.06
Correct at 2-back	.79 (.20)	.68 (.17)	0.050	-0.10	0.00	-0.05
Correct at 3-back	.88 (.27)	.73 (.16)	0.019	-0.14	-0.01	-0.07

Behavioural data are shown in Table 3.3 (above) and Figure 3.2.

Response Accuracy

The bipolar disorder group performed significantly worse than the control group, with fewer correct responses overall. With regard to the different load conditions, the BD group performed significantly worse at 3-back and 2-back and baseline, but not at 1-back. There was no significant difference between the groups with regard to the number of false positive responses. While the overall performance of patients was worse than that of controls, all subjects managed at least one correct response at each load level.

A multilevel mixed-effects linear regression analysis was carried out to investigate the possibility of group, task-level and group*task level effects. Both group and task-level effects were significant ($p<0.0001$), while the group*task level interaction was significant at trend level ($p=0.08$).

Response Times

Response times were significantly slower in the BD group for correct responses, but not false positives. The response time differences for correct responses were significant at all load conditions apart from 2-back, where there was a trend towards a difference (p=0.054).

Covariates

In order to address other possible sources of variance in the data, the analyses were repeated with age, IQ, handedness and gender entered as covariates. When these covariates were added, all the differences in response accuracy and response latency remained significant, except for the difference in correct responses at baseline.

Although other studies have co-varied for subsyndromal symptoms, in an attempt to 'control' for group differences, this is a controversial statistical approach and so was not attempted here. The reasons for this are addressed in detail in the appendices.

Comparison of Behavioural Results to the Literature.

The behavioural results observed here are in general accordance with the N-back literature. Drapier et al[119], who used the same version of the N-Back found that patients performed significantly worse that control and relatives (with no difference between relatives and controls) and that there was a significant group X task interaction. In the only published non-fMRI case control N-back study of bipolar disorder, Harkavy-Friedman et al[188] used the N-Back to compare 30 BD I and 15 BD II patients to 39 controls (neither the N-Back version used, nor the variables compared are reported). In this study, both BD I and BD II patients performed worse than controls; there were no differences between subtypes of bipolar disorder. Adler et al[114], who used the 2-back, also found that patients performed worse that controls (significantly at baseline and at trend level for 2-back). By contrast Monks et al[115], did not find any evidence of a group performance difference on the 2-back version of the N-back (p=0.94). However, given that performance differences appear to become more pronounced at higher memory loads, use of the 2-back only may have reduced the ability to detect performance differences. Finally, Frangou et al[116], did not report any behavioural differences, however their sample (n=7) was underpowered to detect differences.

More generally, the behavioural results are in accordance with what would be expected from what we know from the neuropsychological literature, where the most convincing evidence for neuropsychological dysfunction in bipolar disorder is in the domains of working memory and executive function (see Chapter 1.3).

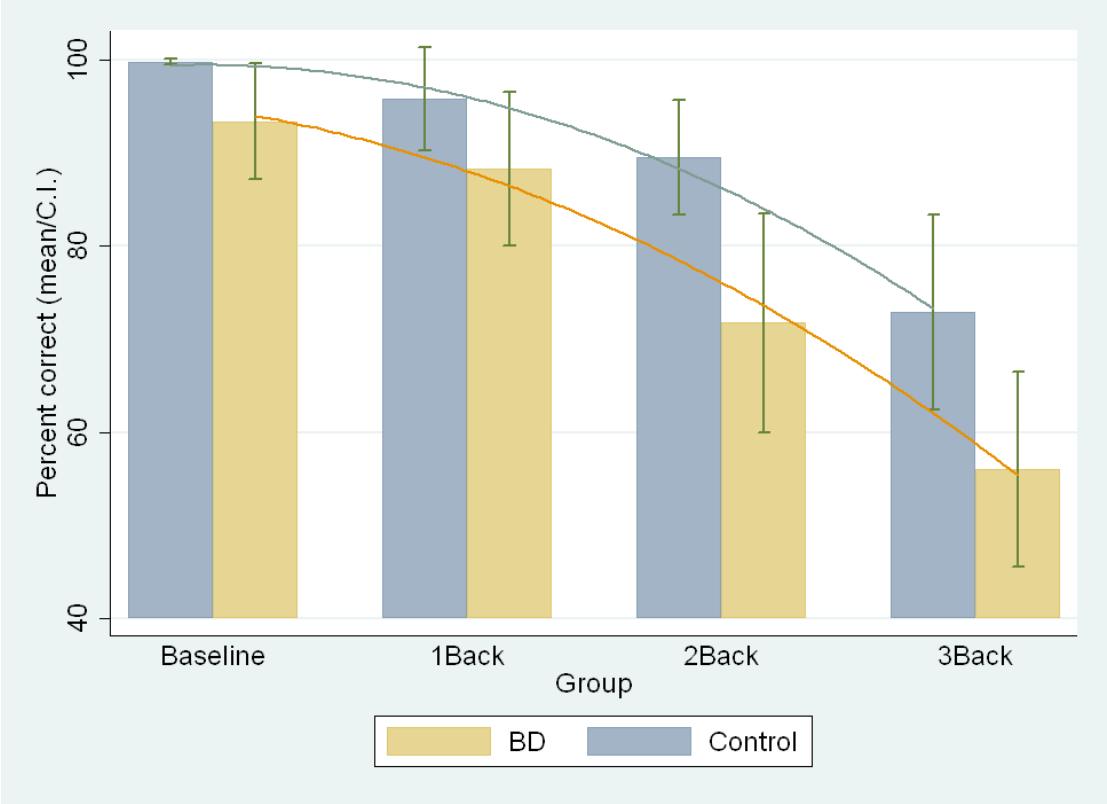

Figure 3.2 Effect of Group and Memory Load on Performance (note, Y axis is abbreviated).

3.3.3. Task Related Activation Networks

Combined BD and control sample (Table 3.4 and Figure 3.3)

T-contrasts were generated to explore the positive and negative effects of the combined (1+2+3 back) active condition on brain activation for: (i) all subjects (BD + control group), (ii) BD patients and (iii) control subjects. For the combined sample, during the active conditions, the following areas showed significantly increased BOLD signal (see Figure 3.3):

- **Bilaterally:** Superior parietal gyrus (extending to angular and supramarginal gyri), middle frontal gyrus, superior frontal gyrus, inferior frontal gyrus, basal ganglia and middle temporal gyrus.
- **Right side only:** midbrain, globus palidus and thalamus.
- **Left side only:** inferior temporal gyrus.

Significantly decreased activation was observed in the following areas

- **Bilateral:** Cingulate gyrus, posterior insula and cuneus.
- **Right side only:** precentral gyrus.
- **Left Side only:** medial frontal gyrus, superior frontal gyrus, superior temporal gyrus, angular gyrus, bilateral, postcentral gyrus fusiform gyrus and superior temporal sulcus.

Both the BD and control group activated qualitatively very similar networks when compared to each other and the combined group. To check whether the same networks were activated in both groups, for each group contrast, Z-scores were extracted using coordinates from the combined sample. Coordinates and Z-scores for the combined sample, as well as the individual BD and control groups can be found in table 3.3.

All of the areas of increased activation from the combined sample were also observed in both the control and the BD groups, at least at trend level ($p<0.01$). Similarly, all areas of decreased activation from the combined sample were also observed in both the control and BD groups (at least at trend level, $p<0.01$), apart from the right precentral

gyrus and left superior temporal gyrus, which were not present in the BD group, even at the trend level.

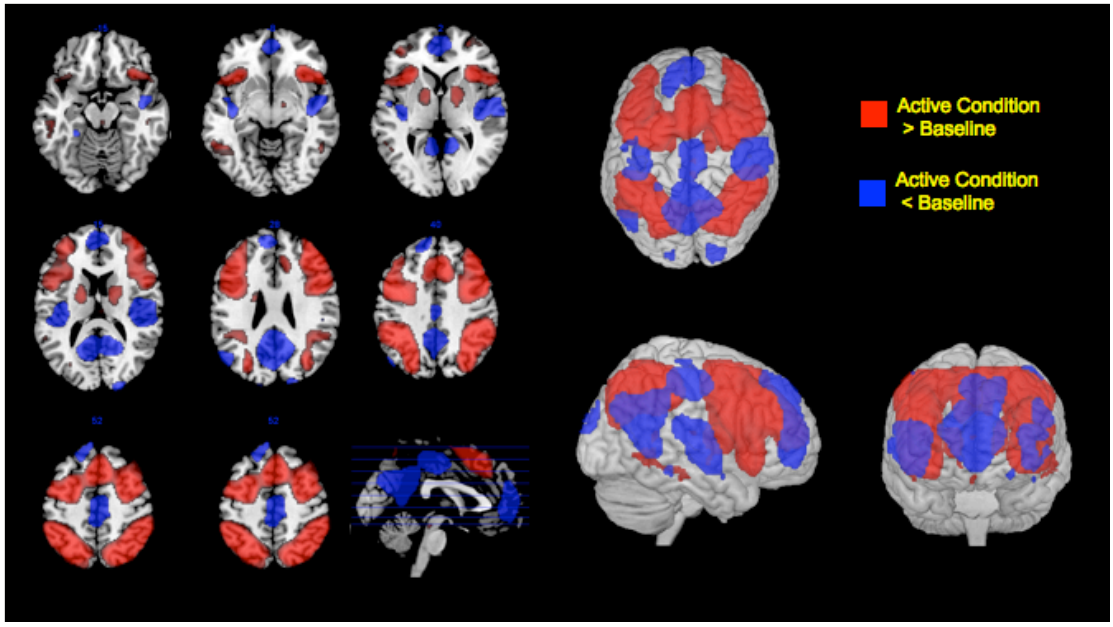

Figure 3.3 Combined Sample. Increased and Reduced Activation for Task Vs Baseline

3.3.4. Effect of Memory Load on Activation

An F-contrast was generated to investigate the main effect of task load on activation. The results of the F-test revealed that activation varied with memory load in a number of regions. In order to investigate the directionality and load specificity of these differences, the following T-contrasts were generated: (i) 3-back > 2-back, (ii) 3-back < 2-back, (iii) 3-back > 1-back, (iv) 3-back < 1-back.

No significant differences were detected between 3-back and 2-back conditions, but differences did exist between 2-back and 1-back conditions. Given the lack of any significant difference between the 2 and 3-back activations, to maximise power, new contrasts were generated to compare the combined 2 and 3 back conditions against the 1-back condition. The results of these contrasts are shown in figure 3.3 and table 3.4.

The following areas demonstrated significantly greater activation in the 2+3 back condition than in the baseline condition (see Table 3.5 and Figure 3.4):

- **Bilateral**: superior frontal gyri, superior parietal gyri (extending to angular and supramarginal gyri) and basal ganglia.
- **Left**: middle/inferior temporal gyrus.
- **Right:** midbrain and thalamus.

Areas that demonstrated significantly reduced activation at 2+3 back relative to baseline were:

- **Bilateral**: posterior insular, cingulate gyri, superior + medial frontal gyri and precentral gyri.
- **Left**: parahippocampal gyrus and superior temporal gyrus.
- **Right**: superior occipital gyrus and cuneus.

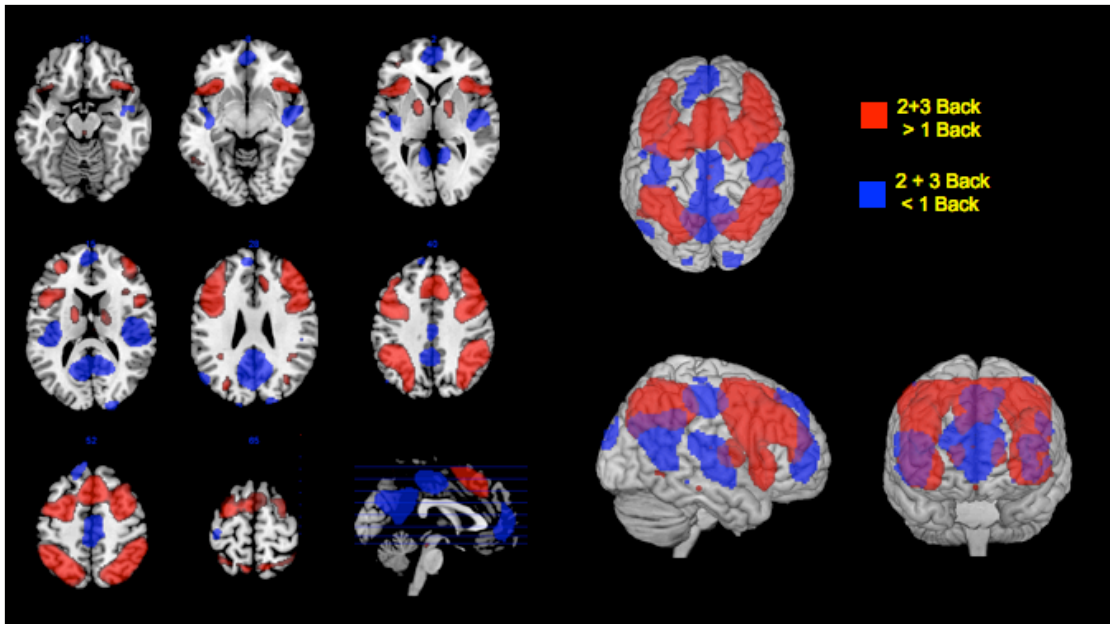

Figure 3.4 Effect of Increased Load on Activation

3.3.5. Group Differences

T-contrasts were generated to investigate whether activation in the two groups differed at either 1-back, 2-back or 3-back levels. In order to maximise power to find differences between groups, as 2 and 3 back loads did not differ significantly in terms of their activation patterns, a further analysis was carried out combing 2 and 3 back levels. Finally, the possibility of a group by task load interaction effect was also investigated using an F-contrast. No areas of significant between-group activation differences or interactions were detected in any of the above comparisons.

3.3.6. Post-Hoc Analysis Excluding Alcohol Dependent Subjects

Four patients had a current co-morbid diagnosis of alcohol dependence and one has a past co-morbid diagnosis of alcohol dependence. While this was just a small proportion of the total patient sample, alcohol dependence is known to have effects on fMRI imaging[189], thus it was possible that this may have affected the results. The analyses were therefore rerun excluding these patients. No significant changes to the results of the previous analyses (either behavioural or fMRI) were found as a result of this reanalysis.

3.3.7. Post-Hoc Analysis of DNB-Control and DB-DNB Differences.

The failure to detect differences between the patient and control groups meant that there were no regions of interest for analysis of differences in the unaffected cotwins of bipolar twins (discordant non bipolar –DNB). There was however a possibility that

there may have been differences between DNB twins and control twins that were not present between the patient and control groups. Given this, I decided to conduct a group analysis comparing unaffected cotwins to control twins. This was done with exactly the same methodology as the patient-control comparison, but using unaffected cotwins rather than patients. Twenty unaffected cotwins were compared with the same set of control twins. As in the previous comparison, the standard activation network reported in Owen et al[190] was present in both groups. There were however, no significant between group differences. There was also a possibility of differences between discordant bipolar and DNB twins. Again, a post-hoc analysis was run to investigate this possibility, but no differences were found.

3.3.8. Post-Hoc Analysis of Differences in Frontal Pole

Drapier et al and Adler et al both reported hyperactivation of the left frontal pole in unaffected relatives of bipolar patients (during 1-back and 2back) and in bipolar patients (at trend level during 1-back only). It was possible that such differences exisited in the present study, but did not survive correction for multiple comparisons. In order to investigate this, the relevant contrasts were rerun with a threshold of $p<0.05$ uncorrected and small volume corrections (8 voxel spheres) were conducted using the coordinates from Drapier et al (-29,59,-7). This procedure is essentially a region of interest analysis on averaged brain images and thus avoids multiple correction adjustment. No differences were found either between bipolar twins and controls or between DNB twins and controls.

Table 3.4 Areas of Activation / Deactivation for Combined Active Conditions vs Baseline.

Z-Scores					MNI Co-ords			Hemi-sphere	Region
Control		BD		Combined	X	Y	Z		
Baseline > 1+2+3 Back									
7.66		7.92		>8.00	-40	-48	46	L	Superior Parietal Gyrus - extending to angular and supramarginal gyri and associate visual cortex (BA 5,7,19,39,40)
7.69		7.81		>8.00	44	-46	44	R	Superior Parietal Gyrus - extending to angular and supramarginal gyri and associate visual cortex (BA 5,7,19,39,40)
7.54		7.98		>8.00	32	2	50	R	Middle Frontal Gyrus (Lateral Premotor – BA 6)
7.50		7.87		>8.00	6	20	50	R	Superior Frontal Gyrus - Extends bilaterally including dorsal anterior cingulate (BA 8,32)
7.17		7.40		8.09	34	24	-4	R	Inferior Prefrontal Gyrus (BA 47, 38)
6.48		6.54		7.67	-32	22	2	L	Inferior Prefrontal Gyrus (BA 47, 38)
6.74		7.20		7.55	-32	2	50	L	Middle Frontal Gyrus (Lateral Premotor - BA6)
4.70	*	4.66		6.51	18	2	8	R	Basal Ganglia
4.33		5.11		6.06	-48	-60	-6	L	Ventral Fusiform Gyrus (BA 37)
4.28		4.53		5.50	-16	4	4	L	Basal Ganglia
3.14	*	4.63		5.38	52	-54	-10	R	Ventral Fusiform Gyrus (BA 37)
3.62	*	3.40	*	4.95	48	-26	-18	R	Middle Temporal gyrus (BA 20/21)
3.34	*	5.11		4.88	-56	-30	-14	L	Middle Temporal gyrus (BA 20/21)
3.25	*	3.26	*	4.59	2	-28	-18	R	Midbrain
2.88	**	3.66	*	4.57	12	-8	-8	R	Basal Ganglia
2.42	**	4.66		4.55	4	-20	14	R	Thalamus
1+2+3 -back < baseline									
7.09		4.41		>8.00	-8	-58	22	L	Posterior Cingulate
6.19		5.00		7.75	10	-56	22	R	Posterior Cingulate
6.12		5.30		>8.00	40	-14	2	R	Posterior Insula
6.20		5.34		7.37	-2	56	10	L	Medial Frontal Gyrus - extending bilaterally
5.38		4.87		7.11	-14	44	50	L	Superior Frontal Gyrus
5.36		4.45		6.66	-38	-18	14	L	Posterior Insula
3.20	*	3.43	*	4.53	-48	-2	-6	L	Superior Temporal Gyrus
5.38		3.38	*	6.24	-48	-74	34	L	Angular Gyrus
4.30		4.44		6.09	16	-96	22	R	Cuneus
4.18		3.19	*	5.24	-42	-26	66	L	Postcentral Gyrus
3.43	*	3.66	*	4.99	-18	-98	24	L	Cuneus
3.89	*	2.75	**	4.77	-26	-40	-16	L	Medial Fusiform Gyrus
3.36	*	NA		4.55	48	-20	62	R	Precentral Gyrus
3.53	*	NA		4.35	-44	2	-18	L	Superior Temporal Sulcus/gyrus

*trend at 0.001 uncorrected **trend at 0.01 uncorrected

Table 3.5 Load Response. Areas of Differential Activation / Deactivation from Baseline for 2+3 Back vs 1-Back Conditions.

Combined Z-Score	MNI Co-ords			Hemi-sphere	Region
	X	Y	Z		
2-Back+3-Back > 1-Back					
>8	6	20	50	R	Superior Frontal Gyrus - Extends bilatterally
>8	30	0	50	R	Middle Frontal Gyrus
>8	42	30	38	R	Middle Frontal Gyrus
>8	-38	-48	46	L	Superior Parietal Gyrus - extending to angular and supramarginal gyri
>8	46	-46	48	R	Superior Parietal Gyrus - extending to angular and supramarginal gyri
>8	32	-62	52	R	Angular Gyrus
7.84	18	-66	60	R	Superior Parietal Gyrus
5.54	-16	-6	10	L	Basal Ganglia
5.41	14	-2	10	R	Basal Ganglia
4.73	-56	-58	-6	L	Middle/Inferior Temporal gyrus
4.51	0	-26	-16	R	Midbrain
4.22	2	-18	14	R	Thalamus
2-Back+3-Back < 1-Back					
>8	50	-28	18	R	Posterior Insula
6.89	42	-14	2	R	Posterior Insula
7.59	-8	-60	24	L	Cingulate Gyrus
6.92	8	-58	22	R	Cingulate Gyrus
7.48	-38	-18	12	L	Posterior Insula
4.84	-46	-2	-6	L	Superior Temporal Gyrus
6.87	0	-14	50	R	Cingulate Gyrus / MFG. Cingulate sulcus.
5.87	-2	-32	54	L	Paracentral Lobule
6.72	26	-94	18	R	Superior Occipital Gyrus
6.38	18	-94	26	R	Cuneus
6.25	0	58	2	M	Superior Frontal Gyrus
5.97	-12	44	52	L	Superior Frontal Gyrus
5.26	-6	54	22	L	Medial Frontal Gyrus (extends bilaterally)
5.58	-54	-68	30	L	Angular Gyrus
5.52	-16	-96	26	L	Superior Parietal Gyrus
4.95	-40	-26	66	L	Precentral Gyrus
4.22	48	-20	62	R	Precentral Gyrus
4.17	-28	-36	-12	L	Parahippocampal Gyrus

3.4. Discussion

The study found evidence of significant performance differences between bipolar disorder patients and matched controls. Network activation and deactivation during the task performance was inline with previous studies[183], for both groups. However, the study was unable to detect any significant activation differences between the two groups. Thus we must reject our primary hypothesis. As no differences in activation were detected between patients and controls, hypotheses two (that the differences seen between patients and controls would also be seen in unaffected cotwins) could not be tested. This is not to say that there could not be differences between unaffected cotwins and controls or between unaffected cotwins and patients. However, subsequent analysis of the data revealed no such differences. As a result, it was not possible to test hypothesis three, that observed differences would be more pronounced in discordant DZ than in discordant MZ pairs.

3.4.1. Strengths and Weaknesses of the Study

Issues Associated With Twin Samples

The present study has a number of inherent limitations, which stem from the use of a twin sample. The first limitation, which applies to any twin study, is that twin studies may not be generalisable to the general population. As discussed in more detail elsewhere (2.9.3), twins are subject to a more complicated perinatal environment than singletons and this may have adverse developmental consequences. As such, caution must be advised in generalising from twin studies to the general population. However, as the study only includes twins and thus compares like for like, this problem should be largely controlled for.

A second limitation associated with twin studies: non-independence of observation in twin pairs, is particularly problematic in the current study due to the combination of twin design and whole brain imaging methods. This is an issue that cannot currently be adequately modelled in standard whole brain imaging software, especially in a sample such as the current one, which involves both within and between-group non-independence. This necessitated adopting a non-standard approach that reduces the sample size available for the initial analysis. This is clearly not an ideal solution, but in

the absence of further development in whole brain imaging designed to address this issue, it was considered to be the best solution available. It is hoped that future iterations of available software packages will include the ability to adequately model such designs (the issue is discussed in more detail in section 2.9.1).

The use of twins also presents issues with regard to recruitment in psychiatric studies. Although the sample size included in the current study is fairly large compared to the existing literature, ideally it would have been larger. The sample described includes all twin pairs, where one or both had a diagnosis of bipolar disorder that could be identified and recruited within a four year period, and who were suitable for and consented to an MRI scan. This recruitment was, of course, subject to limited resources in terms of staff and time, but the primary difficulty stems from the fact that the sample of twins with a diagnosis of bipolar disorder is relatively rare.

Issues Of Sample Size and Power

Despite the above problems of recruiting psychiatric twin samples, the present study had a larger sample size (31 patients compared to 20 in the next largest study) than any published bipolar disorder patient-control study to date and thus should have been suitably powered to detect group differences with similar effect sizes to those from previous studies. There are however, a number of important caveats to this assumption. Apart from sample size, the specific MRI scanner characteristics, specific task, scan length and the analysis method all affect the effective power of a study to find group differences[191]. Thus, direct comparison of power can only easily be made for studies that share all of these characteristics. Samples from different studies also inevitably have different characteristics and different levels of inter-subject variability, which will also affect the power to detect group differences. Further, simply because a prior study has reported a significant group difference, this does not mean that it is has optimal power; beyond the possibility of type I errors, it may be that the study was underpowered to detect more subtle, but important group differences. Nevertheless, according to Thirion et al[126], as a general principle, 20 subjects or more should be used in order to obtain reliable results, while data from studies should ideally be analysed using non-parametric rather than parametric statistics. A weakness of the current study is that, while the current study had well in excess of 20 subjects, the analysis was

conducted using parametric statistics, which may be less sensitive than the non-parametric statistics employed in previous studies.

Patient Heterogeneity

Partly due to the difficulties of recruiting a twin sample of bipolar disorder, the clinical sample used in this study was relatively heterogeneous. Patients were included both with and without a family history of bipolar disorder, with and without a history of psychotic episodes and were recruited both via a wide range of methods. Further, both patients with bipolar I and II subtypes were included (although very few of the latter had an MRI scan, and exclusion of these patients did not alter results). It is possible that recruitment of a more circumscribed phenotype might have resulted in greater power to detect between group differences. However, such an approach would have greatly limited the number of patients that could be recruited.

Medication and Sub-Syndromal Symptoms

Medication and sub-syndromal symptoms represent potential confounds in the current study. Normally, such confounds are offered as possible alternative explanations that for observed group differences, however it is possible that mood stabilising and antipsychotic medication may have a normalising effect on brain function[161] and anatomy. It was not possible to address this within the current study, given that only a few patients were medication free at the time of scanning. Sub-syndromal symptoms are very unlikely to have masked group differences; as it is more likely that they would cause, rather than mask group differences.

Task Related and Scanner Issues

The current study did not identify any between group activation differences, despite the presence of behavioural differences. It is often assumed that group differences in performance will be reflected in BOLD signal changes, but this is not necessarily the case. Further, it is possible that differences in task performance may disguise differences in brain activation between patients and controls. This might be the case if, in a specific area, group membership and performance both correlated with BOLD signal, but in an opposite direction, thus cancelling each other out. On the face of it, this does not seem particularly likely (as it would require considerable spatial overlap of these opposing correlations), but it is not impossible.

In the current study, performance differences were theoretically accounted for by the event related analysis technique, which modelled the BOLD correlates of errors and responses in order to subtract them from the effect of interest – in this case BOLD response during task monitoring. However, poor performance on the N-Back may result in increased frustration, which is difficult to control for, and which was not addressed in the current study. It is possible that the BOLD correlates of unintended and unmodelled confounds such as frustration and anxiety may have reduced power to detect group differences.

One way of dealing with the problem of performance differences between groups is to use a paradigm that is capable of adjusting task difficulty in response to participant performance. However, in a between group study, this introduces further confounds, whereby although performance is equal across groups, one group may be performing a quantitatively different task.

With regard to other similar studies in the literature, only one fMRI study of memory in bipolar disorder has failed to find differences in activation, whilst also reporting behavioural deficits. Monks et al[115] investigated both N-Back and Sternberg tasks in bipolar patients and controls. For the Sternberg task, a group difference (at trend level) was detected for response accuracy, in absence of any activation differences. Here, the authors suggest that the maximum memory load may have been too low to sufficiently challenge the patients cognitive resources - and that it is possible that if the memory load was increased further, activation differences may have been observed, as well as more significant performance differences.

Finally, both the scanner and the scanning parameters used in this study might now be considered sub-optimal. Modern high field strength (3Tesla+) MRI scanners offer the tempting possibility of higher spatial and/or temporal resolutions. Use of such scanners may allow increased sensitivity and specificity for activated voxels and may facilitate the detection of more specific focal differences than were detectible in the current study[192]. Higher resolution may also allow optimisation of the normalisation process.

3.4.2. Comparison with the Literature

The finding of unaltered brain activation during a working memory task is at odds with the published literature. To date, seven papers have been published investigating brain

activation correlated with working memory in bipolar disorder. Four of these used they n-back paradigm and of these, three have reported group differences[119,114,115], while one did not report group differences[116]. The other three papers used the Sternberg task, two of which (analysing the same data, using an emotion version of the task) found differences, while one found no differences. However, despite a majority of studies finding group differences, there has been very little consistency in the findings reported.

It is always difficult to pinpoint what is responsible for disparate findings in imaging studies - as there are wide number of possible causes. Of these, the primary candidates are: differences in the patient and control samples, differences in the task used, difference in the scanners used and differences in the analysis techniques. Comparison with the published n-back literature is most relevant in this discussion. Of the four published papers using the n-back, three[119,115,116] were conducted in the same scanner (at the Maudsley hospital, London) as the present study. The fourth paper, of Adler et al, used a 3 Tesla scanner, a difference that could account for some inter-study difference. The three Maudsley scanner studies all used not only very similar (verbal stimuli) versions of the n-back, but also used the same analysis software, thus inter-study differences are likely due to either sample selection, random sample differences or sample size.

In the case of Frangou et al[116], given the extremely small (7) sample, it is likely that even if differences did exist, there was not enough power to detect them. Drapier et al[119] is the perhaps closest matched to the present study, as it has a comparable sample size (20). Here the most obvious difference was in the sample used; all of the patients were specifically recruited from multiply affected families and all had experienced psychotic or hallucinatory symptoms. In the present study, by contrast, subjects were recruited primarily on the basis of having a diagnosis of bipolar disorder and being part of a twin pair. It is probable that, due to coming from multiply affected families, the sample of Drapier et al[119] had a higher genetic loading for bipolar disorder. In direct contrast (as discussed in the methods chapter), in the present study, our selection criteria may have resulted in a less severely affected sample. Recent studies comparing patients with psychotic and non-psychotic bipolar disorder have suggested that the cognitive deficits reported in bipolar disorder are considerably more severe when the clinical picture includes psychosis[193-195]. It is possible therefore these differences in the

samples used may be reflected in the different fMRI findings. Indeed Drapier et al concede that their 'highly selected sample may not be generalisable to more epidemiological samples of BD'. Monks et al[115] do not provide any information about the psychotic or family psychiatric history of their patients, which suggests these factors were not part of the selection procedure, however in common with Frangou et al[116], this study had a fairly small sample size, which may reduce the generalisability of its results. It is important to note that the observation that highly selected samples may not be generalisable to a wider population is not intended as a criticism. Such differences are simply observed as they are important to note when comparing the results of similar studies. Indeed, selection of more circumscribed clinical samples is an important way of reducing heterogeneity in the data and indeed is arguably an essential approach for advancing our understating of psychiatric disorders. In the case of the current study, such an approach is ruled out because of the difficulty of recruiting a large twin sample with even a relatively broad psychiatric phenotype.

In terms of analysis techniques, the present study used an event related parametric design (analysed in SPM), while the studies above all used proprietary XBAM software developed at the Institute of Psychiatry and Adler et al used proprietary CHIPS software about which little information is available. The XBAM software uses non-parametric statistics that may provide more sensitivity to differences, but the software used does not have the flexibility to allow the more sophisticated modelling that was used in the present study. It is not presently possible to say which of these factors (or what combination) may have resulted in the differences. One solution would be to reanalyse our data using the same analysis methods as Drapier et al[119] (or vice-versa). If, after running both studies thorough the same analyses, between-study differences remained, this would suggest that the different results were due to differences in the samples. If conversely, the results were more similar, it would be reasonable to conclude that the differing analysis techniques were responsible for the prior differences.

Potential Differences in Network Connectivity but Not Network Activation

Although no bipolar-control group differences were detected in the network activated during the n-back task, it is possible that there were differences in connectivity between brain regions that were not detectible in this analysis. In order to address this possibility I conducted further analyses to investigate functional and intrinsic connectivity in our sample; this is detailed in the following chapter.

3.4.3. Conclusion.

The findings from fMRI investigations of working memory in bipolar disorder remain equivocal. Indeed, even when studies have been conducted using the same paradigms and in the same scanners, there has been very little overlap in the results. Clearly then, much more work is needed to establish reliable findings regarding the neural correlates of working memory deficits in bipolar disorder. It is essential that future studies must take full account of the problems encountered to date, specifically those of subject heterogeneity, sample size and analysis techniques. Given the relatively high cost of scanning studies, it may be prudent to more thoroughly examine the issue of the consequences of patient heterogeneity through offline neuropsychological testing before proceeding with further imaging studies.

4. A fMRI Connectivity Analysis of Verbal Working Memory in Bipolar Disorder, in a Twin Sample.

4.1. Introduction

(N.B. The current chapter assumes familiarity with the methods and results of the previous chapter).

In the previous chapter, I investigated the suggestion that bipolar disorder was associated with changes in brain activation during a working memory task. No significant differences were detected between patients and controls, despite differences in behavioural performance. Considering the relatively small behavioural differences detected between groups, it is likely that the differences in brain activation that presumable accompany such behavioural differences were too small to detect, or that they were hidden by the observed differences in performance.

It is possible however, that there are detectable (bipolar-control) group differences in connectivity between brain areas that are not detectable via standard group mapping techniques. That is to say, that the relatively simple subtraction analysis of the BOLD signal may have missed subtle underlying brain dysfunction. Given this possibility, I decided to reanalyse the results, this time investigating the underlying connectivity between key areas involved in the N-Back network.

It is important here to introduce a number of different types of connectivity, in particular: 'anatomical connectivity' 'functional connectivity' and 'effective connectivity'. Anatomical connectivity, as the name suggests, refers to the physical, white matter connections between brain areas, as well as to the neurochemical properties of such connections. Functional connectivity generally refers to the simple correlation between brain activities in two regions and thus may or may not represent a meaningful interaction between two brain regions (for instance activity in two brain areas may correlate due to being driven by that in another brain area). Effective connectivity, by contrast, refers specifically to connectivity between regions that may be identified as representing an influence of one area on another, rather than a simple correlation of activity.

A number of neuroimaging analysis techniques have been developed to investigate effective neural connectivity, including psychophysiological interaction (PPI) analysis, structural equation modelling and dynamic causal modelling (DCM). DCM and structural equation modelling are generally used for model based analysed. PPI,

by contrast can be used to investigate network connectivity in a model free, exploratory manner. Given that no regions of differential BOLD response were detected in the previous analysis, it was not clear what would be an appropriate model for the N-Back. Thus, I used PPI analysis to investigate whether bipolar disorder was associated with any abnormalities in connectivity originating in the key brain areas identified in the previous analysis. The PPI method is explained in more detail in the results section below.

Hypothesis:

Relative to control subjects, patients with a diagnosis of bipolar disorder will exhibit abnormalities of connectivity within the N-back activity network identified in chapter 3.

4.2. Materials and Methods

4.2.1. Sample

This secondary analysis used the same sample as in the chapter 3.

4.2.2. Psychophysiological Interaction (PPI) Analysis.

Brief introduction to PPI

In neuroimaging, psychophysiological interactions are interactions whereby the degree of contribution of one brain area to another brain area is dependent on the experimental context. Friston et al[196] explain psycho-physiological interactions as follows:

'If one were to regress the activity of one region, on the activity of a second region, the slope of this regression would reflect the influence the second area could be exerting over the first. If one then repeated this regression, using data acquired in a different context, then the slope might change. This change in slope is a psycho-physiological interaction.'

For a very detailed discussion of PPI analysis and its interpretation please see Friston et al[196].

In the current study, I identified seven key brain areas implicated in N-Back task performance (from chapter 3). The regions were located in: inferior parietal lobe, middle frontal gyrus, and inferior frontal gyrus (each bilaterally) as well as an area of right medial superior frontal lobe adjacent to the anterior cingulate. The areas are detailed, with their MNI coordinates in Table 4.1. For each of these seven areas, I used PPI analysis to identify any brain areas where connectivity was greater during memory load conditions than during the baseline (vigilance) conditions. In chapter three, no significant differences in BOLD response were detected between 2-back and 3-back. Therefore, to increase the statistical power available, I chose to compare connectivity at the baseline task against connectivity in the combined 2 and 3-back levels (memory load task). The 1-back task was not included as, in the previous analysis, the activated network was significantly different from that in the 2 –back and 3-back levels.

Data Processing and Analysis.

Individual Subject Level Analysis (First Level)

SPM-5 has a built in PPI analysis function. This function extracts time series data from a source brain area and estimates to what extent the functional connectivity between this area and the rest of the brain is modulated by an experimental variable of interest.

Prior to PPI analysis, it is necessary first pre-process the data (individual subject data was processed as described in chapter 3, described under pre-processing and individual subject analysis) and also to extract the time series data, which is done as follows:

1. Load predefined 'effects of interest' contrast for subject and generate statistical map at a low threshold of p=0.9. In this case the 'effects of interest' included baseline, 1-back, 2-back and 3-back conditions, but not response conditions).
2. Select a target voxel (the coordinates of the preselected area of interest)
3. Run a small volume correction to find the local maximum within a spherical search space with a radius of 6 voxels centred on the target voxel. This was done to account for intersubject variability (even with perfect structural registration, the peak activation may not be in the same voxel).
4. Using the SPM VOI function, extract VOI time series for the selected region

Manual extraction of VOI time series data in SPM is very time consuming (using the standard graphical user interface). Therefore, in order to extract time-series data for each region of interest, I automated the process by writing a suitable MATLAB script. Such scripts also reduce the possibility of human error. The script (voiextractor.m) and its dependencies can be found in the appendices.

PPI Analysis.

The PPI calculation itself was done via SPMs built in PPI function. Essentially, the PPI interaction term (or PPI regressor) was computed as the element-by-element product of (i) the target area time series and (ii) a vector which encodes the main effect of memory load condition (coded as -2 for baseline, 1 for 2-back and 1 for 3-

back) convolved with the canonical hemodynamic response function. The PPI regressor was mean corrected and orthoganalised with respect to the main effect of memory load and the target time series.

Brain areas that received inputs from the target area that were stronger during the 2 and 3 back conditions than at baseline were determined by testing for positive slopes of the PPI regressor. This was done by applying a t-contrast that was coded 1 for the PPI regressor and 0 elsewhere. Again, as with VOI extraction, manually processing all subjects would have been extremely time consuming and subject to user error, so again I used MATLAB scripts to automate the process (ppi_masterscript_fk.m and spm_ppi_andrea.m, see appendices).

Group Level Analysis (Second Level)

The individual level contrast images were entered into a full-factorial ANOVA, permitting inferences at the population level. The ANOVA model was specified to include just one factor: group (BD vs control). Contrasts were generated to investigate the average effect of condition and the effect of group. T images for each second level contrast were transformed into statistical parametric maps of the Z statistic. Difference regions were considered statistically significant if they survived family wise error (FWE) correction at the level $p<0.05$.

Table 4.1 Regions Selected for Analysis

	Region	MNI Coordinates
A	Left Inferior Parietal Lobe	-40;-48;46
B	Right Inferior Parietal Lobe	44;-46;44
C	Right Middle Frontal Gyrus	32;02;50
D	Medial Right Superior Frontal Gyrus	06;20;50
E	Left Middle Frontal Gyrus	-28;00;58
F	Right Inferior Frontal Gyrus	34;24;-06
G	Left Inferior Frontal Gyrus	-34;22;-02

4.2.3. Functional Connectivity Analysis.

The PPI analysis showed no (bipolar-control) group differences in connectivity for any of the regions investigated. It remained possible however that differences in

connectivity were expressed similarly during the activation task and the baseline. I therefore decided test this:

Individual Subject Level Analysis (First Level).

The data processing pathway was the same as for the PPI analysis up to and including the extraction of time series data. This analysis is simpler than the PPI analysis however, in that the raw time series data is used rather than the computed PPI regressor. This will be referred to as the VOI regressor.

Brain areas for which activation positively correlated with that of the target region were determined by testing for positive slopes of the VOI regressor.

Group Level Analysis (Second Level)

The individual level contrast images were entered into a full-factorial ANOVA, permitting inferences at the population level. The ANOVA model was specified to include just one factor: group (BD vs control). Contrasts were generated to investigate the average effect of condition and the effect of group. T images for each second level contrast were transformed into statistical parametric maps of the Z statistic. Difference regions were considered statistically significant if they survived family wise error (FWE) correction at the level $p<0.05$.

4.3. PPI Results

4.3.1. All Subject Connectivity Maps

Table 4.2 details for each source region, those areas that showed increased connectivity in the memory load condition relative to baseline. These areas are shown in green in Figure 4.1. All source areas demonstrated increased connectivity in the precuneus (either just right or bilateral). Additionally, increased connectivity was observed in areas corresponding to the left parahippocampal gyrus, lentiform nucleus, temporal lobes, cingulate, postcentral gyri, left cuneus, paracentral lobes, precentral lobes and right middle occipital lobes.

Table 4.2 Areas of Increased Connectivity at 2+3-Back Relative To Baseline, By Source Region

Region	Areas of Increased Connectivity at 2+3 Back Relative to Baseline
Left Inferior Parietal Lobe	Right Precuneus, left parahippocampal gyrus
Right Inferior Parietal Lobe	Right Precuneus, lentiform nucleus
Right Middle Frontal Gyrus	Bilateral Precuneus, Left Temporal Lobe, midbrain
Medial Right Superior Frontal Gyrus	Bilateral Precuneus, Left temporal lobe, Right posterior cingulate
Left Middle Frontal Gyrus	Right Precuneus, Left Precuneus, Left Temporal Lobe, lentiform nucleus, cingulate
Right Inferior Frontal Gyrus	Bilateral Precuneus, right postcentral gyrus, left parahippocampal gyrus, left cuneus
Left Inferior Frontal Gyrus	Bilateral Precuneus, bilateral postcentral gyrus, bilateral paracentral lobe, cingulate, right precentral lobe, inferior parietal lobe, right temporal, right middle occipital

4.3.2. Group Difference Maps

For each VOI, an F-contrast was generated to test for significant bipolar-control group differences (either connectivity in BD < control or visa versa. There were no significant between group (bipolar-control) differences in connectivity for any of the regions investigated.

Reanalysis with separate 2-back and 3-back task levels

A structural equation modelling connectivity analysis of the N-Back by Honey et al[185] reported that different brain networks might be activated during 2-back and 3-back task levels. It was possible therefore, that combining 2 and 3-back task levels confounded the results. To eliminate this possibility, the above analyses were rerun twice, once with a baseline-3back contrast and once with a baseline-2back contrast in place of the initial baseline-(2+3)-back contrast. No group differences were identified in either of these supplementary analyses.

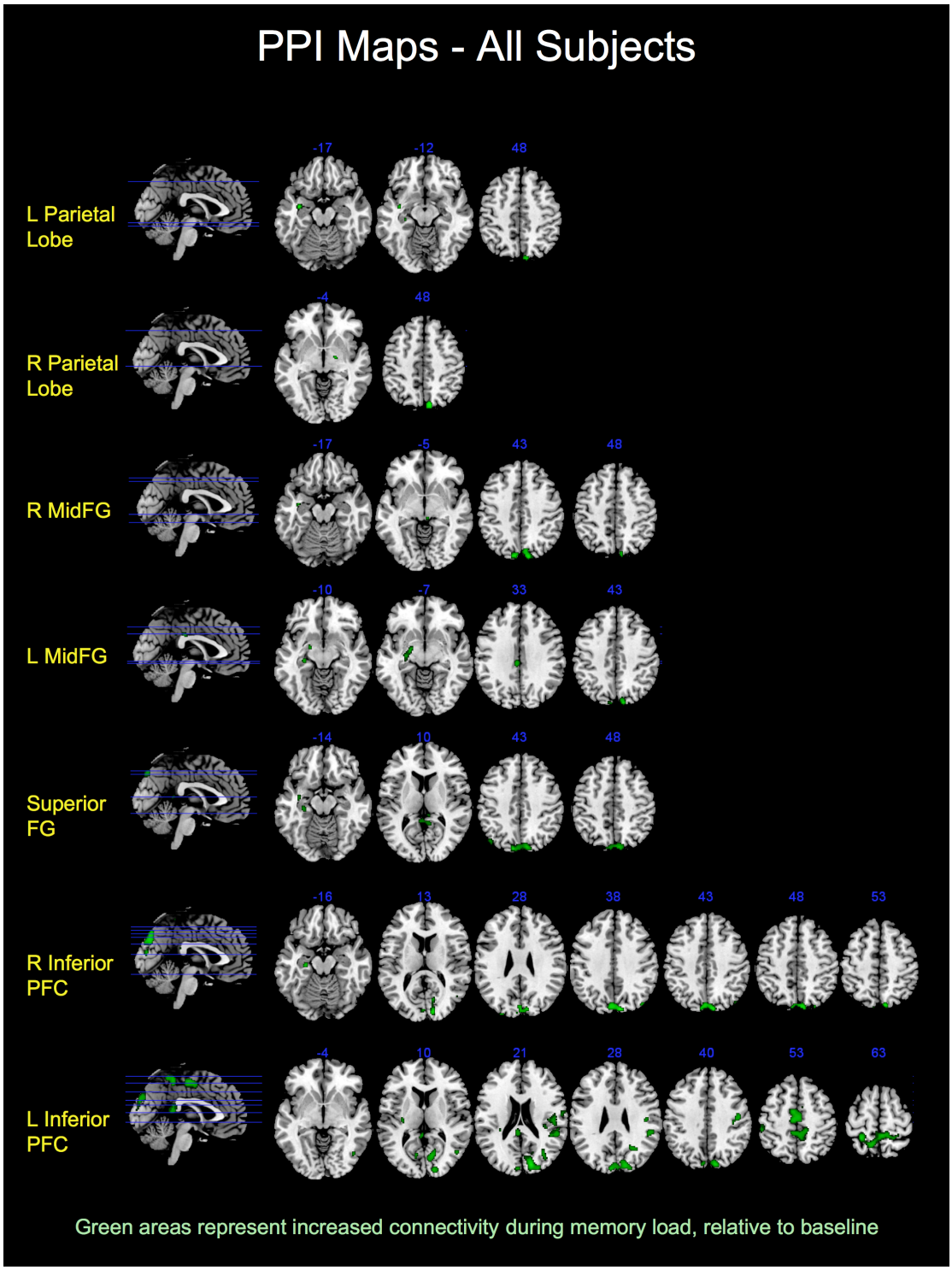

Figure 4.1 PPI Maps. All VOIs, all subjects.

4.4. Functional Connectivity Results

All Subject Correlation Maps

For all VOIs investigated, corrected for Family Wise Error at p=0.05, the functional connectivity brain maps showed significant correlation with the majority of other voxels in the brain. The significance of the observed correlations at the all subject group level was so strong that it was impossible to delineate individual clusters at a normal statistical threshold. Maps were therefore generated a threshold that did allow for better visualisation the areas of strongest correlation ($p<5e-9$).

For every VOI, these areas clearly correspond to the network previously identified in the previous N-Back analysis (chapter 3). Figure 4.2 shows the connectivity maps for all VOIs; on each map, the VOI location is marked by a blue circle. The seven maps are qualitatively very similar, all showing similar correlation maps.

Group Differences

For each VOI, an F-contrast was generated to test for significant bipolar-control group differences (either connectivity in BD < control or visa versa. No differences in functional connectivity were seen between the bipolar and control groups for any source region.

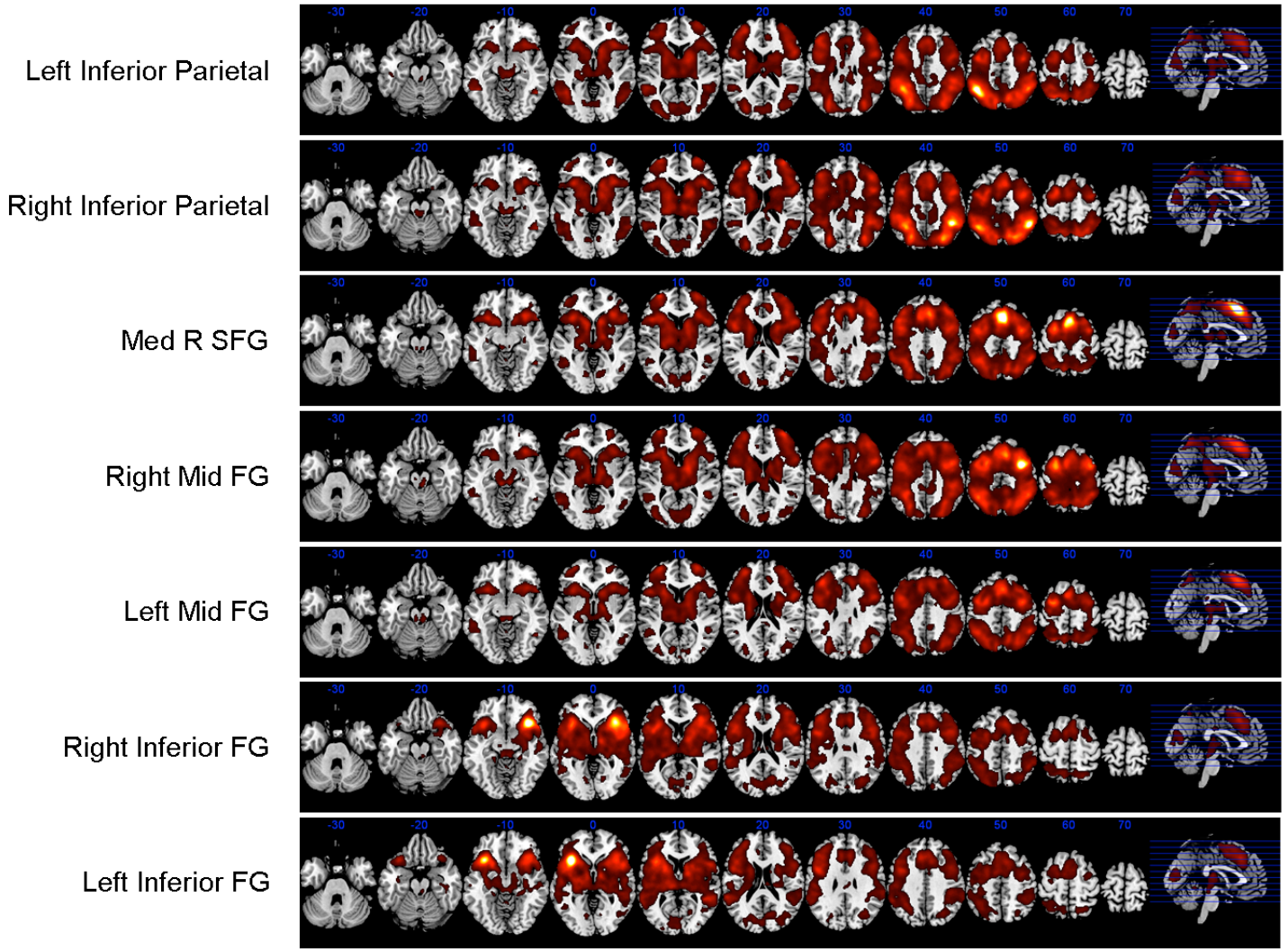

Figure 4.2. Functional Connectivity Analysis. All Subjects Group Maps

4.5. Discussion

This chapter has presented the results of two exploratory analyses investigating the possibility of abnormalities of effective and functional connectivity in bipolar disorder during a working memory task. In neither analysis were any differences found between the neural functioning of patients with bipolar disorder, compared to controls. Thus, we must reject the primary hypothesis: that, relative to control subjects, patients with a diagnosis of bipolar disorder exhibit abnormalities of neural connectivity.

The first analysis used PPI methods to investigate the possibility of task specific connectivity abnormalities in bipolar disorder. For each of the seven regions investigated, a relatively circumscribed set of brain areas showed increased effective connectivity during memory load (compared to vigilance) conditions. Interestingly, for each and every of the seven source regions, increased connectivity was seen in the precuneus during the memory load condition. While the aim of the current study was to investigate bipolar-control differences rather than normal connectivity, this finding is of interest in that it may shed light on the normal role of the precuneus in working memory tasks. Certainly this finding merits further investigation.

The second analysis was a relatively simple investigation of the functional connectivity (time series correlation) between each of the seven ROIs and the rest of the brain. The analysis revealed a widespread network of correlated brain areas, which was very similar to the N-back network seen in chapter 3 and thus to the archetypal N-back network reported by Owen et al[183]. Given the similar activation patterns seen in both analyses, it is reasonable to conclude that the functional connectivity analysis worked as expected.

Given the above it appears that, in terms of detecting areas of both functional and effective connectivity with the chosen source regions, both connectivity analyses were successful. Thus, although it is not possible to rule out the possibility of type II errors in the group analyses (false negative), it is likely that there truly is no *detectable difference* in connectivity between patients with bipolar disorder and controls. The caveat of 'detectible differences' is crucial here. It is important to note that the current study does not rule out differences in the neural functioning of patients with bipolar disorder. More fully, one may accurately conclude only the following: 'Using the

current image acquisition and image analysis techniques, it was not possible to detect abnormalities in the neural connectivity of patients with bipolar disorder during a working memory task'. Indeed, given the extant literature on cognitive function in bipolar disorder, combined with the behavioural differences seen in the current study, it is unlikely that there are *no* neural abnormalities to be observed. It is more likely that given the relatively subtle nature of the neuropsychological/behavioural abnormalities in bipolar disorder, the neural correlates of such abnormalities are similarly subtle.

The strengths and weaknesses of the analyses presented in this chapter are essentially the same as those detailed in chapter three, and so will not be repeated here. It is however, perhaps worth noting that however sophisticated the analysis used, any fMRI study is necessarily limited in that which it can detect. There are a number of reasons for this. Perhaps most importantly, fMRI can only ever measure a proxy of neural activity, namely the BOLD (blood oxygen level dependent) signal, not the neural activity itself[197,198]. Not only is this signal a proxy, but it also reflects mass activity of neurons, rather than the signalling of individual neurons. Furthermore, fMRI studies are limited both in their temporal and spatial resolution and are therefore unable to detect effects that lie within the resolution threshold in either modality. Given the limitations of fMRI imaging, it is not obvious that inefficient or abnormal network activation must always result in differences detectible by current techniques.

Partially as a result of the above, interpretation of changes in activation – and therefore by extension lack of changes, is rarely, if ever, straightforward (see Logothetis[199] for a good summary). Indeed, it is perhaps too easy to provide a plausible interpretation for almost any between group brain difference, in the context of either absence or presence of concomitant behavioural differences. In conclusion then, it appears that, at least within the patient sample studied here, there is no evidence of major neural dysfunction. However, it is possible that subtle dysfunction is present, but undetectable with the neuroimaging methods utilised herewith.

5. White Matter Abnormalities in Bipolar Disorder. A DTI Investigation of Bipolar Disorder in a Twin Sample.

5.1. Introduction

Bipolar disorder appears to be associated with deficits in cognitive function, not only in depressive and manic states, but also in the euthymic state. Evidence from neuroimaging and post-mortem studies suggests that these deficits may be accounted for, at least in part, by white matter pathology. The evidence for this has been discussed earlier in this thesis. The findings from diffusion tensor imaging (DTI) studies in bipolar disorder have also been discussed earlier in the thesis, along with a brief introduction to DTI.

Diffusion Tensor Imaging is a relatively novel extension of traditional MRI techniques that enables researchers to examine diffusion of water within the human brain. The properties of DTI are such that it is uniquely suited to the investigation of white matter. There have been thirteen DTI studies of white matter in bipolar disorder; the majority of these have been region of interest (ROI) studies, with two using both ROI and whole brain analysis, one using only whole brain analysis, two using tractography and one using tract-based skeletal statistics. Due to the varied methodologies, and in particular the different regions of interest in each study, direct comparisons are difficult. However, overall, the evidence indicates that it is possible to detect white matter pathology in bipolar disorder and that this pathology is mainly evident as a reduction in fractional anisotropy.

None of the published studies have used DTI to investigate whether such differences also exist in the unaffected relatives of patients with bipolar disorder. Frazier et al[135] investigated a group of 7 children at risk for bipolar disorder and found evidence of reduced FA (relative to controls) in the superior longitudinal fasciculus. However, given that this study was based on a very small group with mean age of 8.9 years, caution must be given to any interpretation of these findings.

While there is limited evidence of from DTI studies, the structural imaging literature does provide some, albeit conflicting, evidence of a familial and perhaps genetic basis for such white matter abnormalities. For instance, McDonald et al[73] reported an association between genetic risk for bipolar disorder (calculated using a genetic liability scale) and white matter volume reduction in the anterior corpus callosum and bilateral frontal, left temporo-parietal and right parietal regions. This is however in contrast to a similar study by McIntosh et al[74], which did not report such an

association. McIntosh et al[72] also investigated white matter density in patients and their unaffected relatives and reported that while patients with bipolar disorder showed reduced white matter density in the left ALIC, relatives of bipolar patients did not show any differences. However relatives from families with a mixed history of both bipolar disorder and schizophrenia did show reductions in superior frontal subgyral and right medial frontal gyral white matter. It is not clear how to interpret the mixed findings from the studies above, but they offer some tentative evidence that there may be a genetic basis for white matter abnormalities in bipolar disorder.

While evidence from family and twin studies is limited and conflicting, studies of genes associated with both white matter development and bipolar disorder lend support the idea that such differences are genetically mediated. McIntosh et al[200] investigated the effect of a neuregulin 1 (NRG1) variant on white matter. NRG1 is interesting as a number of studies have reported an association between NRG1 and bipolar disorder (especially psychotic bipolar disorder)[201-203]. The study looked at the diffusion tensor imaging and T_1 structural correlates of variation in the rs6994992 single nucleotide polymorphism. 87 normal controls provided T_1 data, of whom a subsample of 43 was scanned with DTI. Based on previous evidence, analysis was restricted to the anterior limb of the internal capsule and frontal subgyral white matter. The structural imaging results showed that relative to subjects homozygous for the T allele (TT), subjects homozygous for the C allele (CC) had significantly higher white matter density in the right ALIC. From the DTI analysis, TT subjects also showed reduced FA of the right ALIC. Similar differences of ALIC white matter density and FA were also observed in the left hemisphere, but were non-significant. Based on these findings and the fact that NRG1 has previously been linked to psychosis, the study's authors conclude that NRG1 'may increase susceptibility to psychosis by altering connections between prefrontal cortex and other brain regions'. Another gene that has been linked to both bipolar disorder[204-206] and brain development is the Disrupted in Schizophrenia (DISC1) gene. Variation in this gene has also been linked to altered FA in the frontal white matter of normal controls[207]. Further, DISC1 has been shown to play an important role in a number of neurodevelopmental processes including neural progenitor proliferation[208] (ablation of neural progenerators has been show interfere to interfere with learning and neural plasticity[209,210]) and neuronal migration[211]. DISC1 has also been shown to interact

with the GSK3B pathway, which is implicated the mechanisms of mood stablisers[208]. Thus although this research is still at an early stage, evidence does appear to be converging, suggesting that genes associated with bipolar disorder may mediate some of the white matter changes that have been observed. However, such research remains preliminary. Indeed, while both DISC1 and NRG1 are candidate genes for bipolar disorder, as noted by Barnett and Smoller[212] in a recent review, neither have yet to be confirmed as susceptibility genes.

Overall, based on the available evidence, it is not yet possible to draw firm conclusions as to whether the observed differences in FA and MD represent a marker of an underlying vulnerability to the disorder, a feature of the disease itself or whether they are due to another confound such as psychoactive medication. Moreover, if such changes do represent an underlying vulnerability to bipolar disorder, it remains unclear how much this is due to genetic as opposed to environmental effects.

The current study aims to address these issues by investigating DTI in identical and non identical twins discordant for bipolar disorder and a matched group of healthy twins. The rationale for this (discussed in more detail earlier) is that if differences found in bipolar patients are also found in their non-bipolar cotwins, this provides evidence of an underlying vulnerability that is not due to disease expression. Furthermore if such differences are greater between discordant fraternal twin pairs than between discordant identical twin pairs, this provides evidence that there is a genetic basis for the difference.

5.1.1. Hypothesis:

Based on the previous literature the following hypotheses were tested:

4. Twins with a diagnosis of bipolar disorder will demonstrate areas of reduced fractional anisotropy relative to control twins. These differences will most likely be located in frontal white matter and/or long range association fibres.

5. Those abnormalities that are seen in twins with a diagnosis of bipolar disorder will also be seen in their non-bipolar cotwins.

6. FA differences between discordant identical twin pairs will be less than FA differences between discordant fraternal twin pairs.

5.2. Materials and Methods

5.2.1. Participants

Participants were drawn from the sample described in chapter 2 (methods). Specific details of sub-sample selection are discussed below, while subject sample characteristics are described in the relevant results sections.

5.2.2. Data Acquisition

Data was collected using a GE Signa LX system (General Electric, Milwaukee, WI, USA), with actively shielded magnetic field gradients (maximum amplitude 40 mT m^{-1}). A standard quadrature birdcage head coil was used for both RF transmission and signal reception. Each volume was acquired using a multi-slice peripherally-gated EPI sequence, optimised for precise measurement of the diffusion tensor in parenchyma, from 60 contiguous near-axial slice locations. Data were collected with a 96x96 matrix size over a 240x240mm field of view, and reconstructed with a 128x128 matrix size to give a final voxel size of 1.875 x 1.875 x 2.5 mm). The echo time was 107 ms while the effective repetition time was 15 R-R intervals (R-R refers to the time between R-wave peaks as measure on a peripheral plethysmograph). The duration of the diffusion encoding gradients was 17.3 ms giving a maximum diffusion weighting of 1300 s mm-2. At each slice location, 7 images were acquired with no diffusion gradients applied, together with 64 diffusion-weighted images in which gradient directions were uniformly distributed in space. Full details are given elsewhere[213]. The DTI acquisition protocol employed at the IoP uses spin-echo EPI based sequences fully optimised for DT-MRI.

5.2.3. Data Analysis.

Non-independence of subjects. A multi-step solution.

The aim of the current study is to identify differences between patients and controls and investigate whether such differences as are seen in the patients are also present in their relatives (in this case their unaffected twins). An inherent difficulty with studies involving twin or family data is that the observations are non-independent. Specifically, due to their shared genetic and environmental influences, family members (especially identical twins) cannot be considered to be independent of each other. The whole-brain imaging packages that are currently available generally use GLM (general linear model) based analysis to investigate differences between groups.

Unfortunately however, such methods include an explicit assumption that observations from each subject are independent. In order to avoid violating this assumption, it was necessary to adopt a three stage analysis approach (the rationale for this is explained in more detail in chapter 2):

- **Stage 1. Preprocessing.** Scan data from all subjects were preprocessed to produce the FA maps needed for the analysis.

- **Stage 2. Group mapping**. In order to examine potential differences between twins with a diagnosis of bipolar disorder and control twins without a diagnosis of bipolar disorder, a group mapping technique was employed. The technique used was adapted from Voxel Based Morphometry (VBM) analysis methods previously developed for structural imaging. Crucially, in this stage, only independent subjects were used. Two groups were selected: a bipolar disorder group (consisting of all BD twins from discordant pairs and one twin from each concordant group) and a control group (consisting of one twin from each pair of healthy volunteer twins).

- **Stage 3. Region of Interest Analysis.** Data were extracted from regions where significant between-group differences were found in the group mapping analysis. The data was then analysed (using a clustering technique to account for non-independence) in order to investigate whether, in these regions, the unaffected cotwins of BD patients differed significantly from either the BD group or the control group.

The three stages are described in more detail below:

Stage 1: Pre-processing

In the current study, the pre-processing stage employed can be divided into two main steps, image generation and image registration. Image generation, in which FA maps are generated from the raw data, was performed using in house software (developed at the IoP). FA and MD maps were calculated using in house software, as per the original definition by Baser and Pierpaoli[214]. Baser and Pierpaoli's definitions allow for the calculation of quantitative data from DTI images. Image registration was performed using the Statistical Parametric Mapping software package (SPM2, Welcome Department of Imaging Neurosciences, University College London).

During the image registration stage, image files from individual subjects were registered to normalise individual differences (for example in overall brain size and brain orientation). The current study utilises a two step registration process, using an analogue of the 'optimised VBM' approach introduced by Good et al.[215] for structural imaging. The optimised registration process involves generating a study specific FA template to which individual FA templates are then registered. This approach reduces the probability of registration errors compared to the standard approach in which a non-specific EPI template is used to register the FA images.

The processing steps involved in the optimised registration process are detailed below:

During the image registration stage, image files from individual subjects were registered to normalise individual differences (for example in overall brain size and brain orientation). The current study utilises a two step registration process, using an analogue of the 'optimised VBM' approach introduced by Good et al.[215] for structural imaging. The optimised registration process involves generating a study specific FA template to which individual FA templates are then registered. This approach reduces the probability of registration errors compared to the standard approach in which a non-specific EPI template is used to register the T_2 weighted (b=0) images.

The processing steps involved in the optimised registration process are detailed below:

1. Registration of each subject's mean T_2-weighted (b=0) image to the standard EPI template provided in SPM. Note, while the b=0 images are T_2 weighted and the template is T_2^* weighted, the image contrast and inherent geometric distortions of echo planar imaging (EPI) based acquisition are similar.

2. Warping parameters derived from step 1 were applied to the corresponding FA image, mapping the FA image into standard space.

3. The normalised FA images from step 2 were averaged and smoothed to create a study specific template.

4. Each subject's FA image was re-registered to the study specific template. The registration parameters determined from this step were also applied to the corresponding MD images, resulting in a set of both FA and MD images in standard space.

5. All the re-registered images were checked to ensure registration had been performed correctly.

6. The re-registered FA images were segmented (using SPM's default a priori tissue probability information), producing maps of the probability of a tissue being either white or grey matter. These segmented images were thresholded at a low (10%) probability to provide a binary mask of white matter. Note: An accurate segmentation was not essential, and a relatively liberal threshold was deliberately used, in order to create a slightly 'over inclusive' mask.

7. Both FA and MD images were mildly smoothed with a 4 X 4 x 4 mm Gaussian kernel to reduce noise and minimize the effects of small residual mis-registrations. (The size of the smoothing kernel chosen at this point affects the sensitivity of the analysis to inter-group differences of differing spatial extents[216]. In the absence of a specific hypothesis about the size of the areas in which we expected to see changes, we chose to use a conservative degree of smoothing, with a kernel size of the same order of magnitude as the width of many white matter tracts).

8. Smoothed FA and MD images were multiplied by the brain masks created in step 6, restricting subsequent analyses to white matter only.

A potential problem with T2* images, such as that used in the template, is that they may have signal dropouts due to susceptibility effects (although visual inspection shows that these are minimal for the SPM template used). The advantages of the T2* template are that it has similar contrast to the T2-weighted images that are collected for the study and that it is EPI based, meaning the it also has similar distortions to the T2 weighted scans from the DTI acquisition. Overall, although not a perfect match, of the templates available, the T2* EPI template is the closest match to the data collected. The first pass registration to the T2* template is also refined using a

second pass FA to FA template registration, for which both image contrast and distortion match between the data and template.

Stage 2. Analysis 1: Statistical Analysis of Group Maps in Independent Subjects.

Subject Selection

In order to avoid violating the ANOVA assumption of independence, subjects were selected such that no two subjects (within and across groups) were members of the same family. Thus in the BD group, all BD twins from discordant pairs were selected, while only one (randomly selected) twin was taken from each concordant pair. In the control group, one twin was selected from each pair; this selection was carried out randomly except when one of the pair met criteria for a psychiatric diagnosis. In such cases, the subjects who did not fulfil criteria for a psychiatric diagnosis were favoured.

Group Mapping Procedure

Investigation of group differences in the maps created in Stage 1 was carried out using the ABAM analysis package, developed at the Institute of Psychiatry, London.

Between-group differences in white matter FA were estimated by fitting a GLM based analysis of covariance (ANCOVA) model at each intracerebral voxel in standard space. i.e.

$$T = a0 + a1V + a2X2 + ... + anXn + e$$

where T is a vector denoting the image value (FA) at a given voxel for each individual in the cohort, V is the independent variable vector (representing group membership), e models the random variation, and the Xn's are covariate vectors representing covariates of no interest if any (Covariates were not entered into the analysis at this stage; instead the effect of covariates was investigated in the extracted data).

The ANCOVA was initially tested with a relatively lenient p-value ($p \leq 0.05$) to detect voxels putatively demonstrating differences between groups. At this stage, only those voxels at which all subjects contribute data were considered, which, along with the

masking procedure (steps 6 and 8, above), restricted the analysis to core white matter regions, reducing the search volume (and thus the number of comparisons made) and also avoided testing at the grey/white interfaces, where the high grey/white contrast of FA images exacerbates any edge effects.

Next, a search was run for spatially contiguous clusters of voxels within those highlighted in the previous step. The 'mass' (the sum of suprathreshold voxel statistics of which it comprises) of each cluster was tested for significance. Permutation based testing, implemented in the XBAM package (developed at The Institute of Psychiatry, London, UK http://www.brainmap.co.uk/), was used to assess statistical significance at both the voxel and cluster levels. For more details of permutation testing please see Bullmore et al[217].

At the cluster level, rather than set a single a priori p-value below at which findings were significant, a calculation was performed to establish, for a range of p-values, the number of clusters which would be expected by chance alone. The statistical threshold for cluster significance was then set such that the expected number of false positive clusters by chance alone would be less than one. In the case of the current analysis, a cluster significance threshold of 0.0025 was used.

Assumptions of smoothness and local permutation testing.

An underlying assumption in the investigation of cluster level effects that all regions will equally 'smooth' and can therefore be treated equivalently (from a statistical point of view). It is known, however[218], that the variance of FA values in the brain depends upon the FA values themselves and, while our analysis was restricted to core white matter regions where this is relatively uniform, the effects of physiological and MR noise may still vary slightly from region to region.

Stage 3. Analysis 2. Region of Interest Analysis

Using the regions identified in stage 2, for each subject (including the cotwins not include in the group mapping analysis), the mean FA and MD values were extracted from each region. Data were then entered into STATA (version 10) for further analysis. This analysis consisted of two steps.

Step 1. Analysis 2a. A new subsample was selected. This sample consisted of all twin pairs discordant for bipolar disorder and all control twins. The sample was divided into three groups:

1. Twins with bipolar disorder from discordant pairs (DB)
2. Twins without bipolar disorder from discordant pairs (DNB)
3. Control twins

Regression analysis was then used to investigate whether, in the regions previously identified, DNB twins differed significantly in their FA or MD values from either their DB cotwins or control twins.

Step 2. Analysis 2b. The DB and DNB groups were further subdivided by zygosity so that any differences identified in analysis 2a could be further examined. If there was a greater difference in FA or MD between DB and DNB twins in DZ than in MZ twins, this would be indicative of a genetic basis for the observed differences.

5.3. Results

(Please note, data was generated and analysed using the raw output from the DTI images, where FA was scaled from 0-1000, rather than 0-1. Thus, FA value on graphs, as well as coefficents from regression analyses should be interpreted accordingly (e.g. FA of 500 on the graphs corresponds to 0.5 on the standard 0-1 scale)). Tabulated means and standard deviations have been transformed into the more standard 0-1 scale.)

5.3.1. Analysis 1. Statistical Analysis of Group Maps in Independent Subjects.

Sample characteristics.

Following re-registration (step four of pre-processing), all images were checked manually. At this stage, one subject (a twin with a diagnosis of BD II from a discordant pair) had to be excluded, as the images could not be successfully registered to the optimised template. This was most likely due the presence of unusually large first and third ventricles. As a result of this, it was not possible to use exactly the same sample that was used in the earlier fMRI analyses.

The BD group thus consisted of 30 patients (10 from concordant pairs and 20 from discordant pairs). The control group consisted of 24 subjects. Demographic and mood characteristics for the two groups are summarised in

Table 5.2. There were no significant differences between groups for gender, ethnicity, handedness, parental social class, age, IQ, or years of education. As would be expected, the BD group did, however, have significantly higher mania and depression scores than the control group.

Psychiatric Diagnoses

Psychiatirc diagnoses and current comorbid conditions are shown in Table 5.1. In the BD group, 29 participants had a diagnosis of BD-I and 1 participant had a diagnosis of BD-II. Of the BD-I participants, four had a comorbid diagnosis of alcohol dependence (one with cannabis dependence), three had a comorbid diagnosis of panic disorder (two with agoraphobia), one had a comorbid diagnosis of generalised anxiety disorder and one had a diagnosis of OCD. In the control group, one participant met criteria for panic disorder.

Table 5.1 Analysis 1. Frequencies of Current Comorbid Conditions (in Bipolar Patients) and Diagnoses + Comorbid conditions (in controls).

Comorbid Conditions (BPAD) or Primary Diagnosis + Comorbid Conditions (Control)	Frequency	
	BD	Control
Alcohol Dependence + Cannabis Abuse	1	-
Alcohol Dependence	3	-
GAD	1	-
OCD	1	-
Panic Disorder w Agoraphobia	2	1
Panic w/o agoraphobia + GAD	1	-

Medication

Of the 30 participants in the BD group, 21 were taking mood stabilisers, of whom 4 were on lithium monotherapy, 4 were on non-lithium mood stabilising monotherapy, 7 were also taking antipsychotics and 6 were also taking antidepressants. Of the remaining, 5 patients were taking no medication, 2 patients were taking only antidepressants, one was taking both an antipsychotic and an antidepressant and one was taking an antidepressant and a benzodiazepine. No controls were taking psychoactive medications.

Table 5.2 Analysis 1. Demographic and Mood Characteristics.

	GROUP		Sig	Test	C.I.		Coef.
	BD	CONTROL			Low	High	
Demographics							
N	30.0	24.0					
Gender (% male)	40.0	20.8	0.132	chi2			
Ethnicity (% white cauc, non white cauc)	93.3	87.5	0.858	FE			
Handedness (% left, right, mixed)	80,13.3,6.7	87.5,4.2,8.3	0.562	FE			
PSC (% I, II, III, IV, V, UE)	10,30,36.7,13.3,10,0	16.7,33.3,33.3,4.2,8.3,4.2	0.767	FE			
Age (sd)	40.9 (13.7)	35.5 (11.5)	0.129	reg	-6.204	0.812	-2.696
IQ (sd)	112.9 (12.7)	114.3 (11.8)	0.688	reg	-2.693	4.051	0.679
Years of education (sd)	15.5 (3.1)	15.5 (2.7)	0.925	reg	-0.761	0.836	0.038
Diagnois: BD I, BD II	29, 1	NA					
Age of onset of mania/hypomania (sd)	26.6 (10.8)	NA					
Years since onset (sd)	14.3 (11.3)	NA					
Mood							
HAM-D (sd)	6.4 (7.8)	0.8 (1.4)	**0.001**	reg	-4.411213	-1.176	-2.794
YMRS (sd)	2.1 (3.3)	0.3 (0.7)	**0.013**	reg	-1.608	-0.197	-0.902

chi2: Chi Squared, FE: Fishers Exact, reg: regression, PSC=parental social class (groups I-V, unemployed). HAM-D: Hamilton Depression Scale. YMRS: Young Mania Rating Scale

Between group FA differences

Relative to controls, patients had significantly lower FA in three clusters, two in the corpus callosum and one in a region identified as the left superior longitudinal fasciculus (regions were identified by reference to a standardised atlas of white matter[219]).

The largest of these areas consisted of bilateral genu and forceps minor, extending laterally to include bilateral internal capsule and also caudally along left body of CC for about 2/3 of its length (i.e. until the middle of the cortex along the anterior-posterior axis) at which point it extended dorsally to include a section of the corona radiata. The second corpus callosum cluster originated in the left anterior side of the splenium, extending laterally to include a posterior section of inferior longitudinal fasciculus (ILF) / inferior frontal occipital fasciculus. The cluster in the superior longitudinal fasciculus (SLF) originated near the ILF and extended rostrally, approximately half way along the length of the SLF. For brevity, these regions will be referred to as the genu, splenium and SLF regions respectively.

There were no areas where FA was significantly higher in patients relative to controls.

Table 5.3 shows the mean FA values for each of the regions, by group. Figure 5.1 shows the FA values for each subject, by group and region. Figure 5.2 and Figure 5.3 show the regions of difference (in a multi-slice view and a 3D rending respectively).

Table 5.3 Analysis 1. Characteristics of regions showing significant patient-control differences.

Region	Size	Peak (talaraich)			Probability	Mean FA (S.D)			
	(voxels)	x	y	z		BD		Control	
Splenium	307	-25	-48	10	0.00107	0.473	(0.028)	0.504	(0.038)
SLF	228	-31	-17	32	0.00135	0.361	(0.036)	0.395	(0.040)
Genu	1167	-13	39	-6	0.00018	0.405	(0.035)	0.433	(0.038)

Effects of demographic and mood variables on FA

Regression analyses were run to investigate whether FA in the three regions identified was significantly associated with IQ, years of education, age, years since onset (in

patients only), gender, handedness or current mood state. Due to the number of comparisons, and thus the increased risk of type I errors, only results with a p value of less than 0.05 are discussed. The results are shown in Table 5.5.

There were significant inverse relationships between age and FA in all three regions. The relationship between age and FA for the three regions is illustrated in Figure 5.4. It was possible that these relationships may have been due to (non-significant) group differences in age. As the regions were defined on the basis that they were different in FA, the fact that the groups were (randomly or not) different in age, could result in a spurious correlation, thus the analyses were rerun within each group. When this was done only the relationship in the genu remained significant, in both controls ($p=0.037$) and patients ($p=0.006$) (Table 5.4). This indicates that the age-FA relationships in the splenium and SLF may have been due to the small group differences in age. It is however, also possible that the loss of significance may have been due to a loss of power due to smaller sample sizes. If one compares the regression coefficients for the overall comparison vs. those of the within group compression, they are similar, lending weight to the later explanation (Table 5.4).

As it is suggested that bipolar disorder may be, at least in part, a neurodevelopmental disorder, it is possible that this would be reflected in abnormal development of white matter over time. It is also possible that the medication taken by patients with bipolar disorder may have long term effects on brain morphometry. It is thus possible that the relationship between age and FA may differ between groups. In order to test for this, a group X age interaction was added to the analysis. There were no significant group X age interactions on FA, in any area. It should be noted that the current, cross sectional study is not ideal for investigation of developmental effects. In order to rigorously test hypotheses including the effects of age, it is necessary to employ longitudinal designs, imaging the brain at multiple points in an individual's lifespan.

There were also significant inverse relationships between Hamilton Depression Scale (Ham-D) scores and FA in both corpus callosum regions, but not the SLF. This finding of a possible relationship between Ham-D scores and FA was intriguing. However, it was possible that this finding was being driven by the group differences in both FA scores and Ham-D scores (in a manner analogous to the FA-age relationship). The analysis was therefore repeated, but this time for patients and

controls separately, the results are shown in Table 5.4 and Figure 5.6. Regression analyses revealed that there were significant negative correlations between Ham-D scores and FA, for all regions, but only in control subjects. Further regression analyses, with an additional term modelling a group by depression interaction, confirmed that, for all areas, there were significant group x depression interactions (splenium p=0.001, SLF p=0.040, genu p=0.021) on FA, which when accounted for, removed the significant HamD-FA relationship in overall sample (splenium p=528, SLF p=0.819, genu p=0.421).

From investigating the graphs of the within group depression-FA relationsips however, it appears that the significant negative correlations seen in the control subjects could represent false positives and should be treated with caution. Such caution should be exercised for two reasons. The first reason for caution is simply that such a relationship was not hypothesised and was unexpected. The second reason is the distribution of the data points. In all three regions, the data are very unequally distributed over a small range, with the large majority of data points clustered at a Ham-D score of 0 (baseline) or 1, with few data points being 3 or greater. It is therefore possible that the observed relationship is due to outliers in the data. Furthermore, such a relationship seems theoretically unlikely, given that Ham-D scores are likely to vary on a day to day basis while FA is presumably rather more stable. The range of the data in control subjects is also small, representing only a small fraction of the possible range of the Ham-D. This is problematic as the Hamilton Depression scale was not designed to capture reliable and sensitive data within such a small range.

There are a number of statistical approaches to dealing with the possible distorting effect of outliers. Firstly, one can drop the outliers and see how this affects the significance of the results. Such outliers can be identified statistically by using (for example) 'Cook's D', which is general measure of the influence of a data point, being a combined measure of the leverage and residual vales of a data point. The standard criterion for assessing whether a data point may represent an outlier is whether it's Cook's D value is greater than 4/n (n being number of subjects). Secondly, one can use alternative statistical tests that are more robust to the effects of outliers. One such example is the Spearman correlation analysis, which does not assume a linear

relationship between variables, being sensitive to any monotonic relationship. The Spearman correlation is equivalent to a Pearson's correlation (or the coefficient of a linear regression) in which the data is transformed from a continuous to a ranked scale prior to analysis. The Spearman correlation is thus fairly robust to the effects of outlying data points. Both approached were attempted here.

Outlier data points were identified as described above. Two data points representing the same subject were dropped from the analysis for all three brain regions, while for the SLF and splenium a further data point was dropped. These dropped data points are highlighted in yellow in Figure 5.6 and consist of both subjects with Ham-D scores of 5 (all regions) and one subject with a Ham-D score of 0. When the regression was rerun excluding these potential outliers, the relationship remained significant in splenium ($p=0.012$) and SLF ($p=0.039$) but not genu ($p=0.062$). The data was also tested with the Spearman correlation analysis (without dropping outliers). Here, the correlation remained significant for splenium ($p=0.004$) and genu ($p=0.019$) but not SLF ($p=0.090$).

It appears therefore, that the relationship between depression scores and FA measures cannot be ascribed simply to outlying data. This leaves two possibilities: firstly that the relationship is genuine and secondly that the relationship is a false positive arising from statistical chance. If the former, this would represent evidence that FA changes were significantly related to even very minor changes in ratings on the Hamilton depression scale. This, if validated, would be even more interesting given the group interaction with this relationship, whereby a FA-depression relationship was seen in controls but not patients.

However, it is difficult to explain what such a relationship might represent, if real. Of the nine controls that had a Hamilton score of over zero, five had a score of just one, two had a score of two and two had a score of five. To obtain a score of one, subjects scored the lowest possible level on just one item (for instance feeling that in the last week they had let someone down, but not having any other symptoms from the considerable symptom list probed by the Ham-D (mood, insomnia, agitation, problems with work, retardation, anxiety, somatic issues, libido, hypocondriasis, weight loss, paranoia etc). When the potential outliers were dropped from the analysis, no subject scored more than two on the Ham-D, yet the relationship

remained significant in the SLF and splenium. It appears highly improbable that a real relationship could be represented within such a small range of data on a scale such as the Ham-D. Nevertheless, investigation of this possible relationship in independent samples would be desirable to confirm/disconfirm this finding.

Table 5.4 Significance and Correlations of FA-Age Relationship, by Region

Area	Overall		BD		Control	
	Coef.	Sig	Coef.	Sig	Coef.	Sig
SLF	-0.94	0.033	-0.74	0.180	-0.68	0.329
Splenium	-0.87	0.024	-0.63	0.088	-0.66	0.383
Genu	-1.51	0.000	-1.45	0.006	-1.28	0.037

Table 5.5 Analysis 1. Correlations between FA and demographic and mood variables.

		CC - Splenium					SLF					CC - Genu				
		coef	sig	CI		R^2	coef	sig	CI		R^2	coef	sig	CI		R^2
				low	high				low	high				low	high	
FA	**IQ**	0.40	0.338	-0.43	1.22	0.00	0.12	0.797	-0.83	1.08	0.00	0.25	0.584	-0.65	1.15	0.02
	Years Ed	1.27	0.469	-2.23	4.78	0.01	0.28	0.890	-3.76	4.32	0.00	1.88	0.324	-1.90	5.65	0.25
	Age	-0.87	0.024	-1.61	-0.12	0.09	-0.94	0.033	-1.80	-0.08	0.08	-1.51	0.000	-2.25	-0.77	0.09
	Ham-D	-1.77	0.024	-3.29	-0.24	0.10	-1.55	0.086	-3.32	0.23	0.06	-1.83	0.032	-3.49	-0.16	0.00
	Young-M	-0.48	0.805	-4.32	3.37	0.00	1.11	0.613	-3.27	5.48	0.01	-0.71	0.734	-4.88	3.46	0.03
	Years since onset*	-0.34	0.476	-1.30	0.62	0.02	-0.22	0.721	-1.49	1.04	0.00	-0.90	0.130	-2.08	0.28	0.08
	Gender	18.61	0.082	-2.42	39.64	0.06	21.17	0.084	-2.97	45.30	0.06	14.75	0.205	-8.32	37.81	0.03
	Handedness		0.254			0.05		0.071			0.06		0.072			0.06

Abbreviations: CC - corpus callosum, SLF - superior longitudinal fasciculus, Ham-D - Hamilton depression scale, Young-M - Young mania rating scale
*Regression run using only DB group

Table 5.6 Analysis 1. Correlations Between FA and Hamilton Depression, By Group

	CC - Splenium					SLF					CC - Genu				
	coef	sig	CI		R^2	coef	sig	CI		R^2	coef	sig	CI		R^2
			low	high				low	high				low	high	
Patients	-0.47	0.512	-1.92	0.98	0.02	-0.21	0.820	-2.10	1.68	0.00	-0.71	0.434	-2.53	1.12	0.02
Controls	-15.92	0.002	-3.45	0.00	-3.45	-11.87	0.040	-23.17	-0.57	0.18	-13.22	0.015	-23.63	-2.81	0.02

Abbreviations: CC - corpus callosum, SLF - superior longitudinal fasciculus

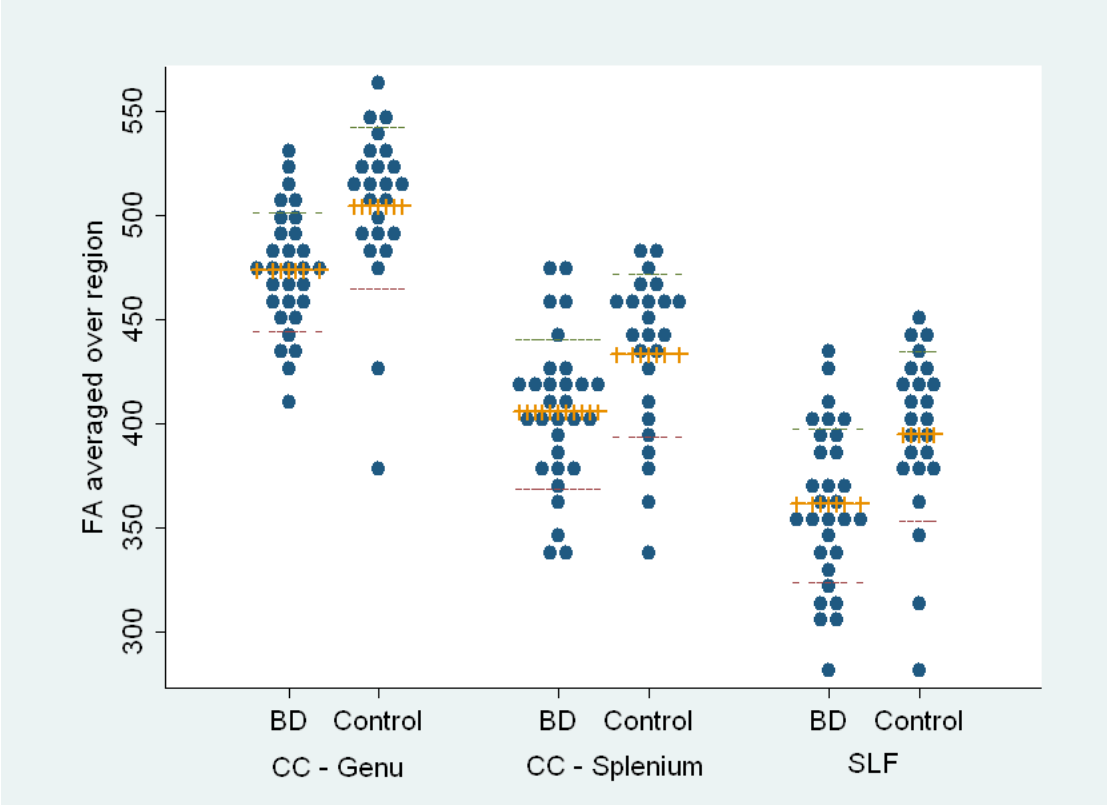

Figure 5.1 Analysis 1: Patient-Control Differences in Mean FA, by Region and Group.

(Note: in this and subsequent graphs, orange crosses represent group means, while dashed lines represent one standard deviation above and below the mean)

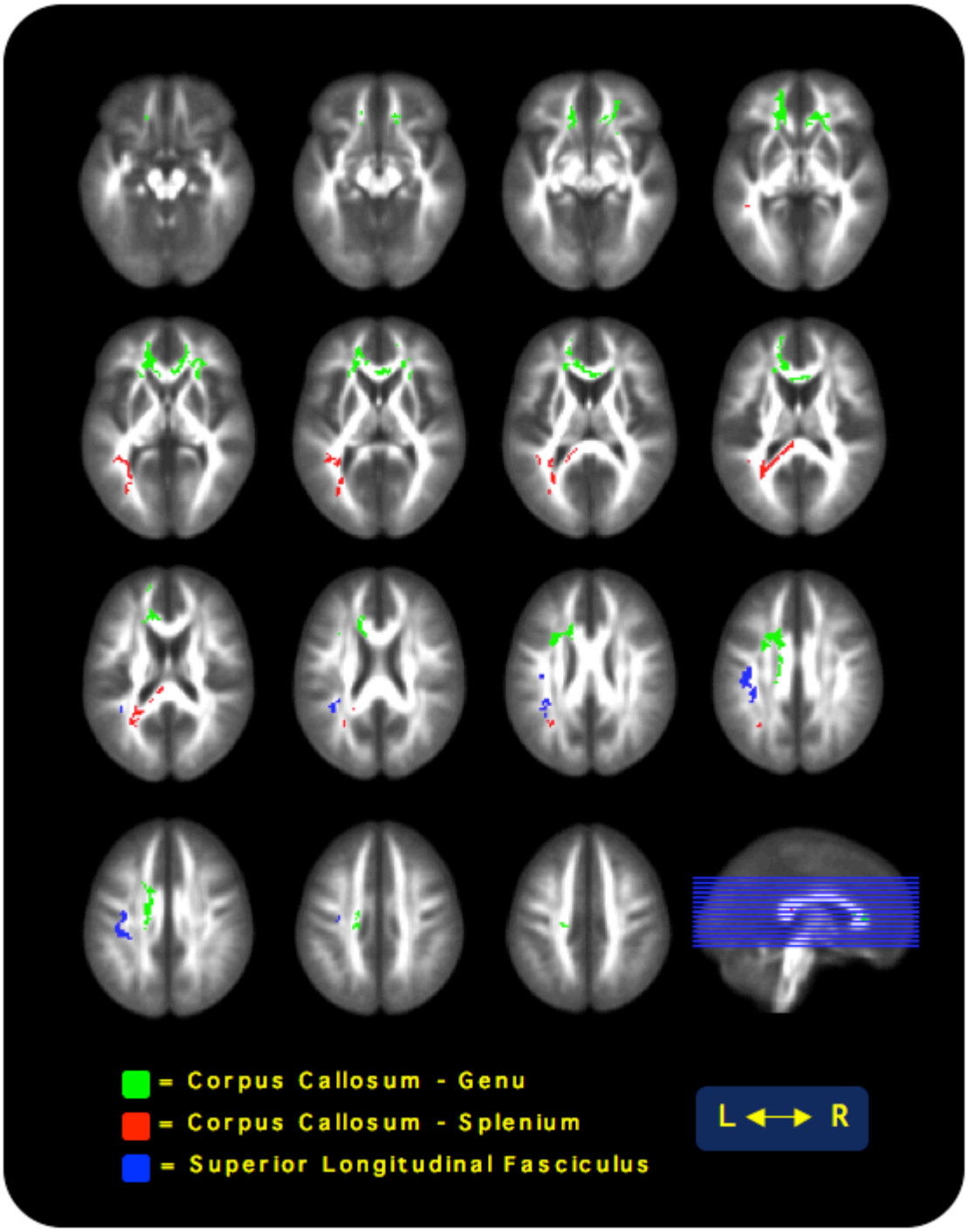

Figure 5.2 Analysis 1. Multi-slice representation of FA differences between patients and controls

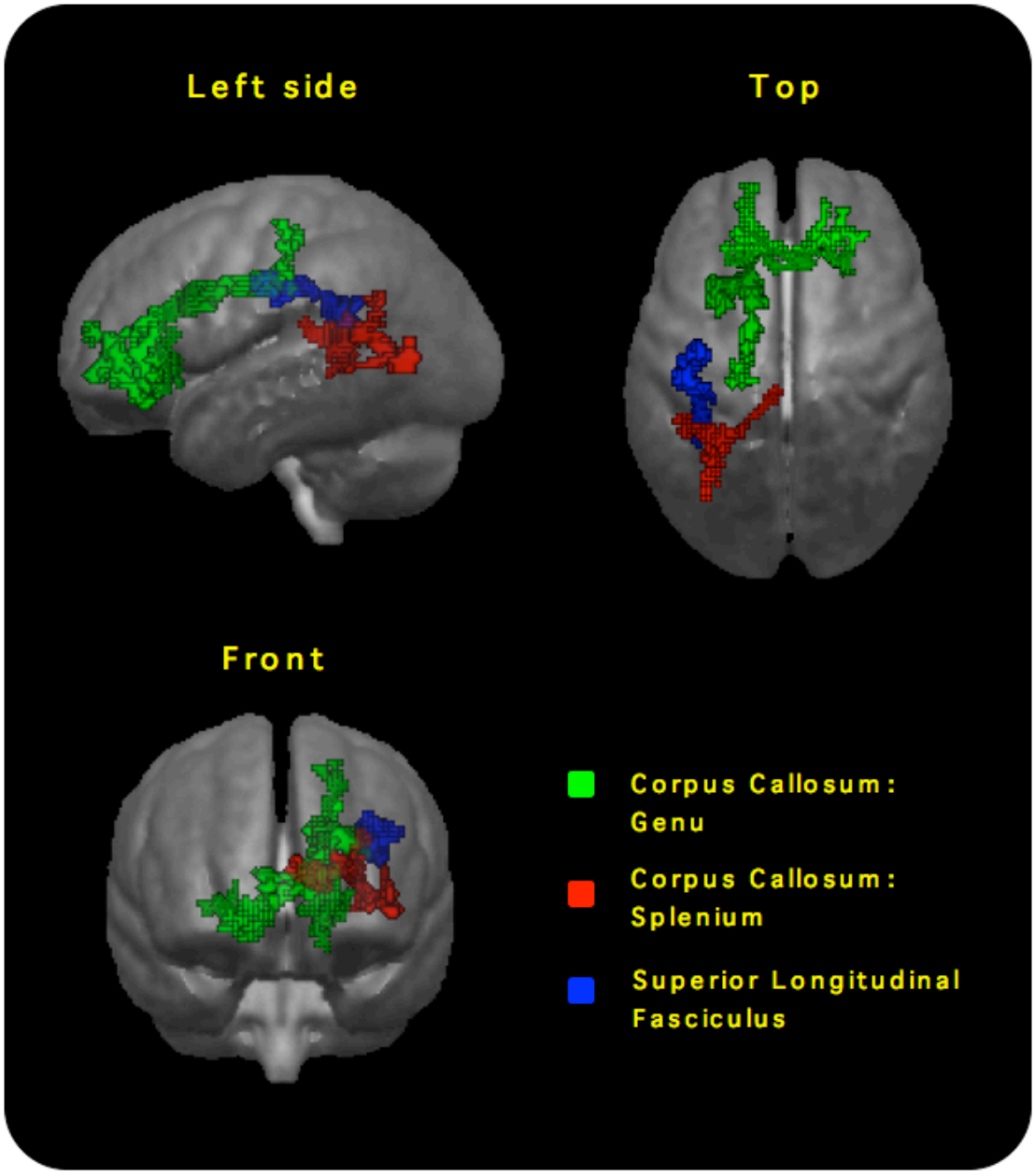

Figure 5.3 Analysis 1. 3D rendering of FA differences between patients and controls.

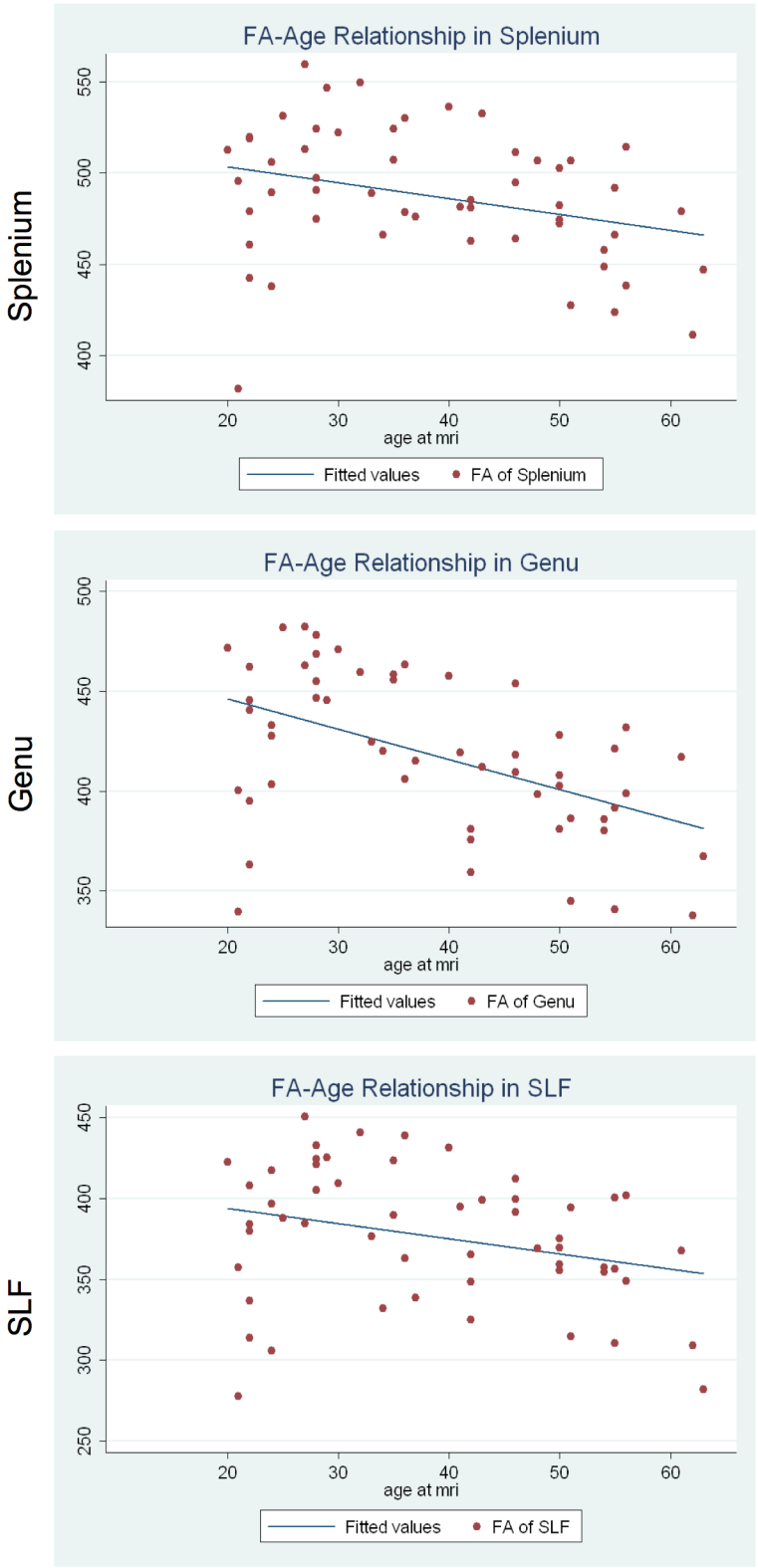

Figure 5.4. Analysis 1. FA-Age Relationship, by Region

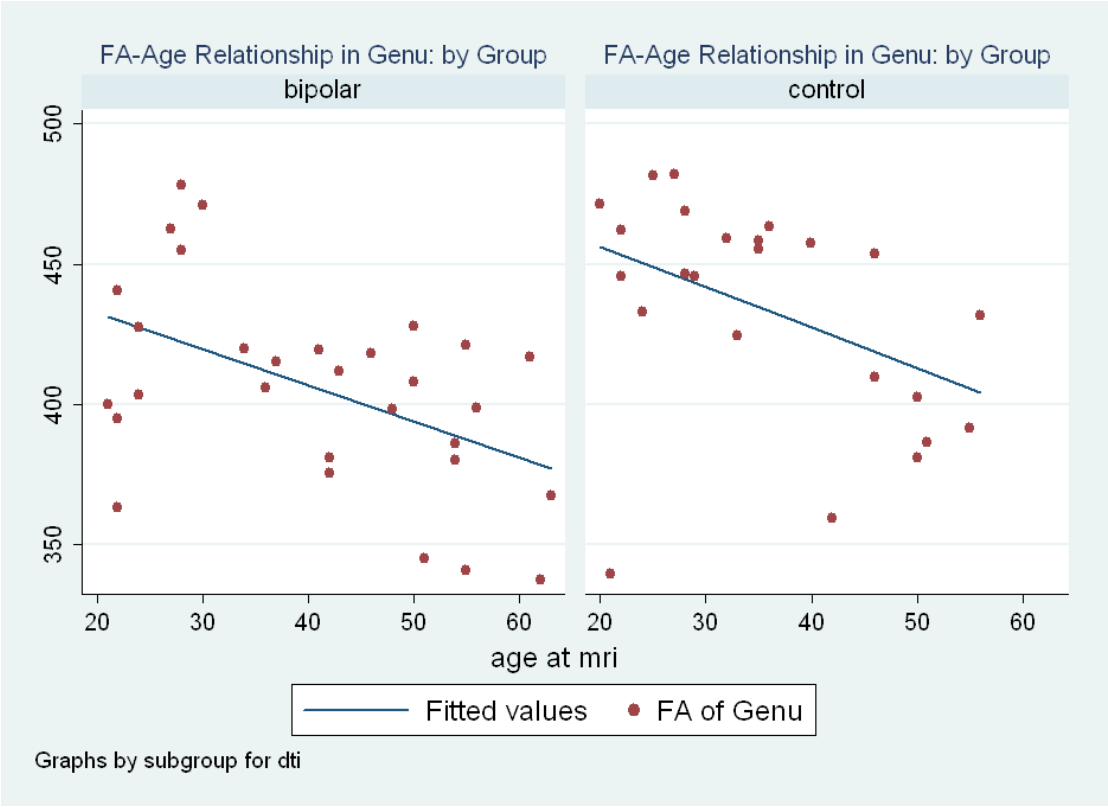

Figure 5.5. Analysis 1. Relationship between FA and Age, in Genu, by Group

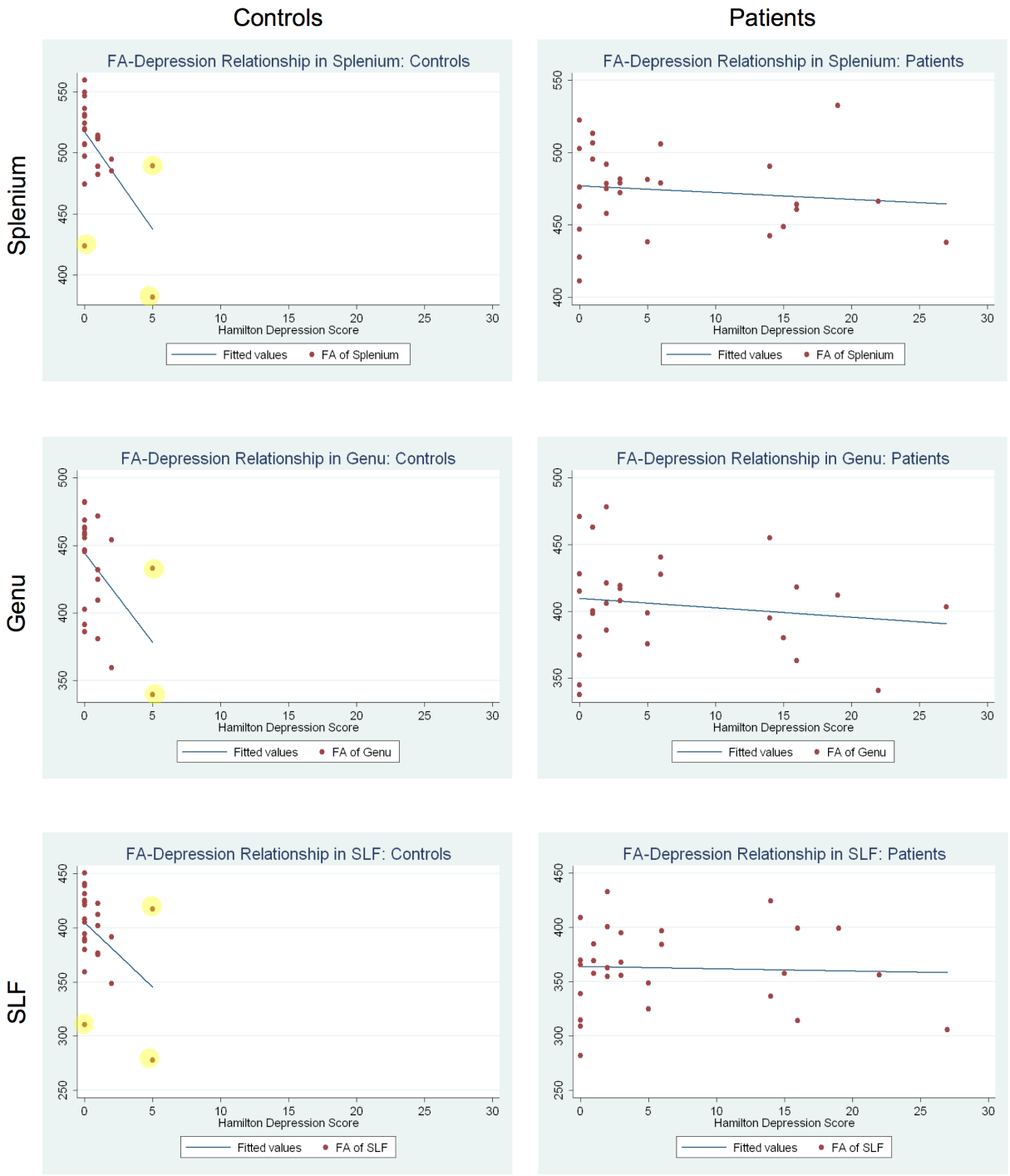

Figure 5.6. Analysis 1. Correlations Between Depression and FA, by Group and Region

5.3.2. Analysis 2a: Three Group, Region of Interest.

Sample characteristics and group differences

The sample consisted of all discordant twin pairs where scans were available for both twins, as well as all control twins. Subjects were divided into three groups for analysis: twins with bipolar disorder from discordant pairs (DB), unaffected co-twins from discordant pairs (DNB) and control twins.

Comparisons of demographic and mood variables were made for discordant bipolar DB vs DNB and DNB vs. control groups. Demographic and mood characteristics, as well as between group test results for the three groups are summarised in Table 5.10.

There were no significant differences between groups for gender, ethnicity, handedness, parental social class, age, IQ, or years of education. The DB group had significantly higher depression, but not mania scores than the DNB group. There were no differences in mood scores between the DNB and control groups.

Psychiatric Diagnoses

Psychiatric diagnoses and current comorbid conditions are shown in Table 5.7. In the DB group, 23 participants had a diagnosis of BD-I and one participant had a diagnosis of BD-II. Of the BD-I participants, one had comorbid alcohol dependence, two had comorbid panic disorder and one had comorbid OCD. Two further participants had a previous history of panic disorder and one had a previous history of alcohol and cannabis dependence. The participant with BD-II had no comorbid conditions.

In the DNB group, 3 participants had a diagnosis of major depressive disorder (one with comorbid panic disorder), one of OCD with comorbid panic disorder, one of panic disorder, one of alcohol dependence and one had a past single depressive episode with comorbid alcohol and cannabis dependence. The remaining 13 had no psychiatric diagnosis.

In the control group, two volunteers had a diagnosis of panic disorder, one had a diagnosis of agoraphobia, two had a past single depressive episode and four patients had a diagnosis of alcohol dependence. One control had a past history of cocaine and

cannabis use. Here it should be noted that the SCAN diagnostic interview is extremely sensitive to any drug/alcohol use.

Table 5.7 Analysis 2a. Frequencies of Current Comorbid Conditions (in Bipolar Patients) and Diagnoses + Comorbid conditions (in controls).

Comorbid Conditions (BPAD) or Primary Diagnosis + Comorbid Conditions (DNB / Control)	Frequency		
	Discordant Bipolar (DB)	Discordant Non Bipolar (DNB)	Control
Alcohol Dependence	1	1	4
OCD	1	1	
OCD w Panic Disorder		1	
Single Depressive Episode with Alcohol and Cannabis Dependence		1	
Major Depressive Disorder		2	
Major Depressive Disorder with Comorbid Panic Disorder w Agoraphobia		1	
Panic Disorder w/o Agoraphobia		1	2
Panic Disorder w Agoraphobia	2		
Agoraphobia			1

Medication

Of the 21 patients, 13 were taking mood stabilisers, of which 3 were taking lithium monotherapy, 3 were taking non-lithium mood stabilising monotherapy, 2 were also taking antipsychotics and 5 were also taking antidepressants. Of the remaining patients, 2 were taking only antidepressants, 1 was taking only antipsychotics, 1 was taking an antidepressant plus an antipsychotic and 1 was taking an antidepressant plus a beta blocker. In the discordant-non-bipolar group one person was taking an antidepressant, one person was taking a hypnotic, one person was taking thyroxin and one person was taking tamoxifen[15]. None of the others were taking medication. In the control group, no participants were taking medication except one taking lisinopril and two were taking thyroxine.

Between Group FA and MD Differences

For every subject, FA and MD values were extracted for each of the three regions identified in analysis 1. DB-DNB and DNB-Control group difference comparisons were conducted for each region. The results of the comparisons are shown in Table

[15] Patient had been taking tamoxifen for 3 years prior to scan, as a recurrence preventative following a sucessful operation for breast cancer.

5.11. FA and MD values each group (by region) are shown in Figure 5.7 and Figure 5.8.

FA

FA was significantly lower in the DNB than the control group in the splenium. There were no other significant differences in the DB and DNB groups relative to controls, and no significant differences between the DB and DNB groups in any region.

MD

There were no significant differences between in either DNB-DB or DNB-control comparisons. In the genu, at trend level, MD was higher in the DNB group than the control group.

Adjustment of results for effect of covariates.

The significance of the above differences remained unchanged when age, gender, handedness and IQ were added as covariates.

Effects of demographic and mood variables on FA and MD.

The relationships between demographic/mood variables and FA and MD were examined using linear regression, adjusted for clustering in the data. Due to the number of comparisons, only differences significant at $p<0.05$ will be discussed. Table 5.8 details all comparisons.

FA

In the splenium, FA was negatively correlated with depression scores (Figure 5.9). In the genu, FA was positively correlated with IQ (Figure 5.12) and negatively correlated with age (Figure 5.10) and depression scores (Figure 5.14).

As appeared to be the case in analysis 1, it was possible that the negative relationship between FA and HamD scores (seen in the splenium and genu) was due to the significant group differences in both FA and HamD scores. Therefore the analysis was rerun for each group. In these subsequent analyses, there was no significant FA-HamD relationship, for any group or area. In Figure 5.9 and Figure 5.14, the scatter plots of HamD against FA are broken down by group, and it is possible to see how the group differences in FA and HamD scores may be generating a spurious relationship.

The FA-IQ and FA-Age relationships were also decomposed into subgroups. When decomposed (Figure 5.13), the FA-IQ relationship in the genu was only significant for the DB ($p=0.019$) group, not for the DNB group ($p=0.551$) or controls ($p=-0.191$). By comparison, when the age-FA relationship in the genu was decomposed by group (Figure 5.11), there was a significant relationship in the control group ($p=0.006$), but not the DB ($p=0.135$) and DNB ($p=0.553$) groups.

There was also a significant relationship between handedness status and FA of the genu. This difference was significant for right-left but not right-mixed or mixed-left comparisons, with right-handers having significantly higher FA than left-handers. Mean FA was 0.420 (s.d. 0.004), 0.396 (s.d. 0.008), 0.412 (s.d. 0.002) for the right, left and mixed groups respectively.

MD

There were no significant effects of for any demographic or mood variable on MD.

Table 5.8 Analysis 2a. Relationship between FA/MD and demographic/mood variables. With Regression Statistics.

		CC - Splenium					SLF					CC - Genu				
		coef	sig	CI		R^2	coef	sig	CI		R^2	coef	sig	CI		R^2
				low	high				low	high				low	high	
FA	**IQ**	0.46	0.248	-0.33	1.25	0.02	1.12	0.462	-1.93	4.18	0.01	0.61	0.038	0.04	1.19	0.03
	Years Ed	0.82	0.566	-2.03	3.67	0.00	1.12	0.462	-1.93	4.18	0.01	2.26	0.090	-0.37	4.89	0.03
	Age	-0.41	0.326	-1.23	0.42	0.02	-0.35	0.420	-1.23	0.52	0.01	-1.05	0.004	-1.76	-0.34	0.14
	Ham-D	-1.57	0.021	-2.89	-0.25	0.04	-1.40	0.137	-3.25	0.46	0.03	-1.86	0.012	-3.29	-0.43	0.07
	Young-M	0.16	0.918	-2.94	3.25	0.00	2.23	0.186	-1.12	5.58	0.02	-0.08	0.947	-2.59	2.43	0.00
	Years since onset*	-0.09	0.849	-1.12	0.93	0.00	0.58	0.522	-1.29	2.46	0.02	-0.91	0.335	-2.83	1.01	0.06
	Gender	15.14	0.104	-3.25	33.52	0.03	5.57	0.647	-18.79	29.93	0.00	0.06	0.994	-17.81	17.94	0.00
	Handedness		0.180			0.04		0.154			0.04		0.012			0.06
	Right-Left											-24.22	0.007	-41.54	-6.89	
	Right-Mixed											-7.71	0.634	-39.94	24.52	
	Left-Mixed											16.51	0.340	-17.81	50.82	
MD	**IQ**	0.25	0.844	-2.25	2.75	0.00	-0.70	0.333	-2.14	0.74	0.01	-0.12	0.872	-1.55	1.32	0.00
	Years Ed	0.30	0.951	-9.52	10.12	0.00	-1.70	0.391	-5.64	2.24	0.00	-2.38	0.329	-7.22	2.46	0.01
	Age	1.47	0.326	-1.51	4.46	0.02	1.34	0.128	-0.40	3.07	0.05	1.35	0.129	-0.41	3.10	0.04
	Ham-D	4.90	0.086	-0.72	10.51	0.03	-0.34	0.712	-2.21	1.52	0.00	1.34	0.226	-0.85	3.52	0.01
	Young-M	0.49	0.908	-7.89	8.87	0.00	-3.56	0.054	-7.18	0.06	0.01	-2.19	0.285	-6.25	1.87	0.00
	Years since onset*	-1.79	0.376	-5.86	2.28	0.03	0.34	0.626	-1.06	1.73	0.01	-0.32	0.741	-2.27	1.64	0.00
	Gender	3.14	0.919	-58.16	64.44	0.00	21.05	0.090	-3.38	45.48	0.02	26.82	0.065	-1.75	55.40	0.02
	Handedness		0.913			0.00		0.060			0.02		0.980			0.00

Abbreviations: CC - corpus callosum, SLF - superior longitudinal fasciculus, Ham-D - Hamilton depression scale, Young-M - Young mania rating scale
*Regression run using only DB group

Table 5.9. Analysis 2a. Relationship Between HamD and FA, by Group. With Regression Statistics.

	CC - Splenium					CC - Genu				
	coef	sig	CI		R^2	coef	sig	CI		R^2
			low	high				low	high	
DB	-0.96	0.246	-2.64	0.72	0.09	-1.47	0.144	-3.49	0.55	0.12
DNB	-3.17	0.279	-9.14	2.79	0.03	-0.39	0.889	-6.18	5.40	0.00
Controls	-2.99	0.473	-11.44	5.47	0.02	-4.71	0.181	-11.77	2.34	0.05

Abbreviations: CC - corpus callosum, SLF - superior longitudinal fasciculus

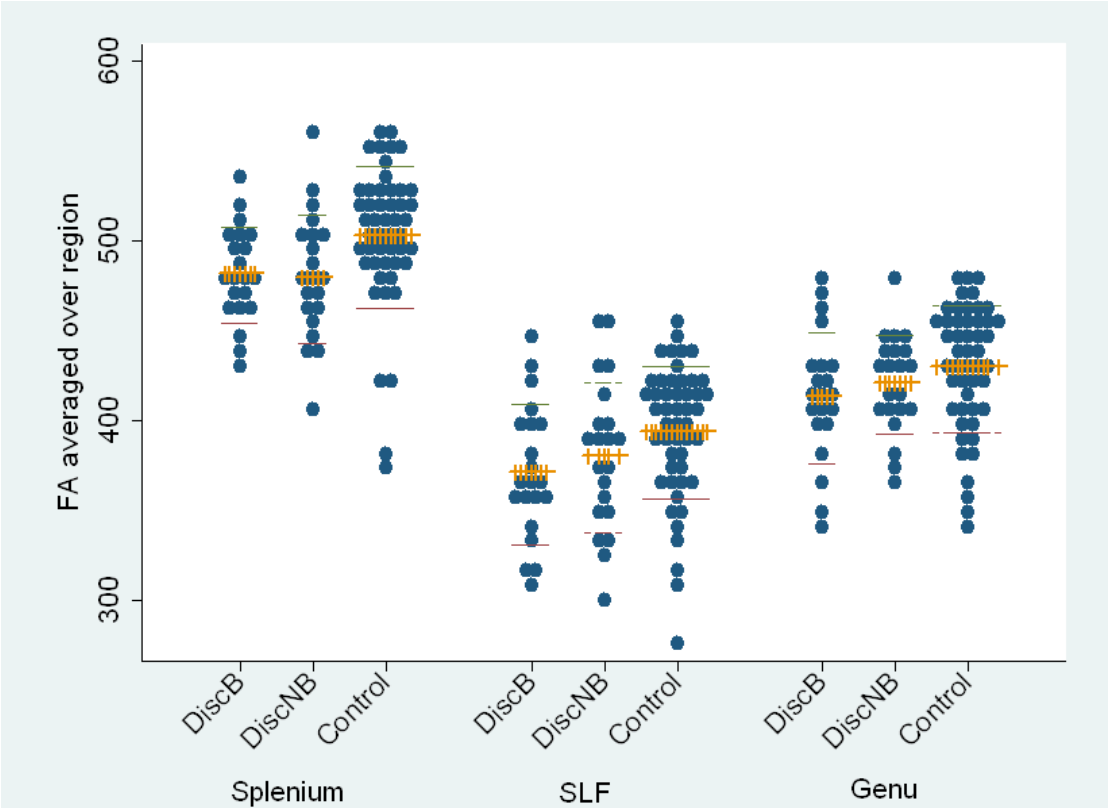

Figure 5.7 Analysis 2a. Mean FA by group and region

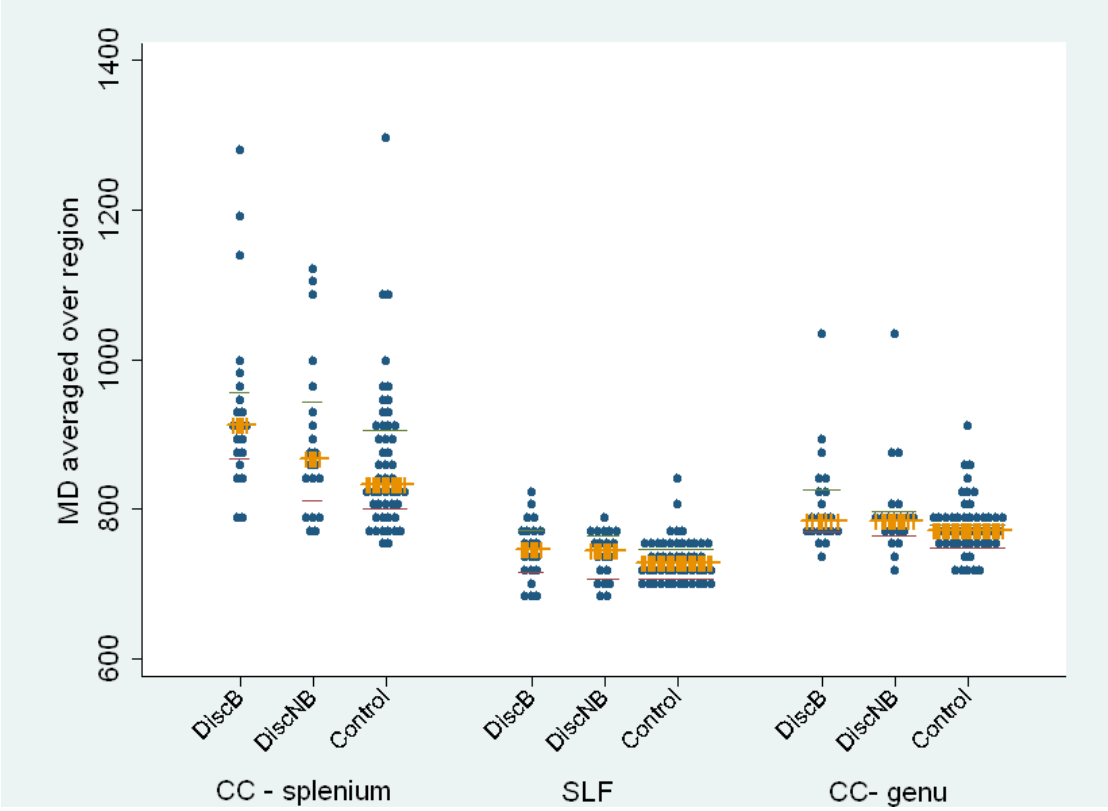

Figure 5.8 Analysis 2a. Mean MD by group and region

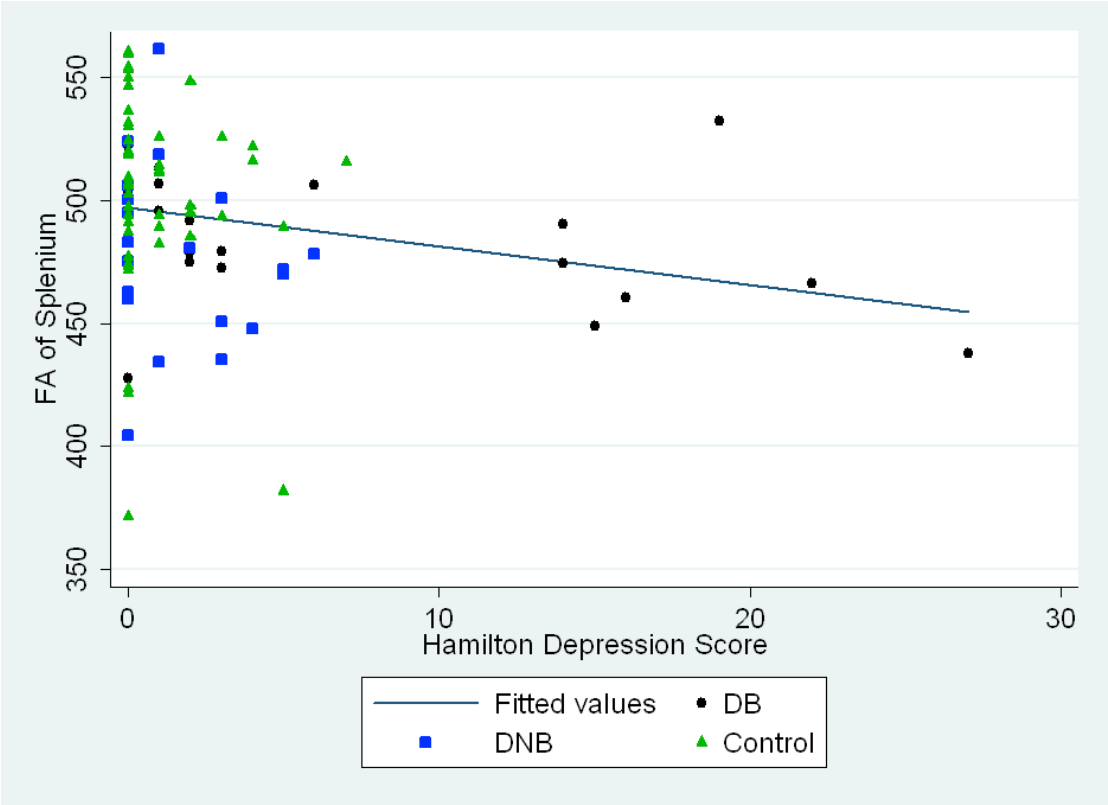

Figure 5.9. Analysis 2a. Relationship between FA and HamD in the Splenium

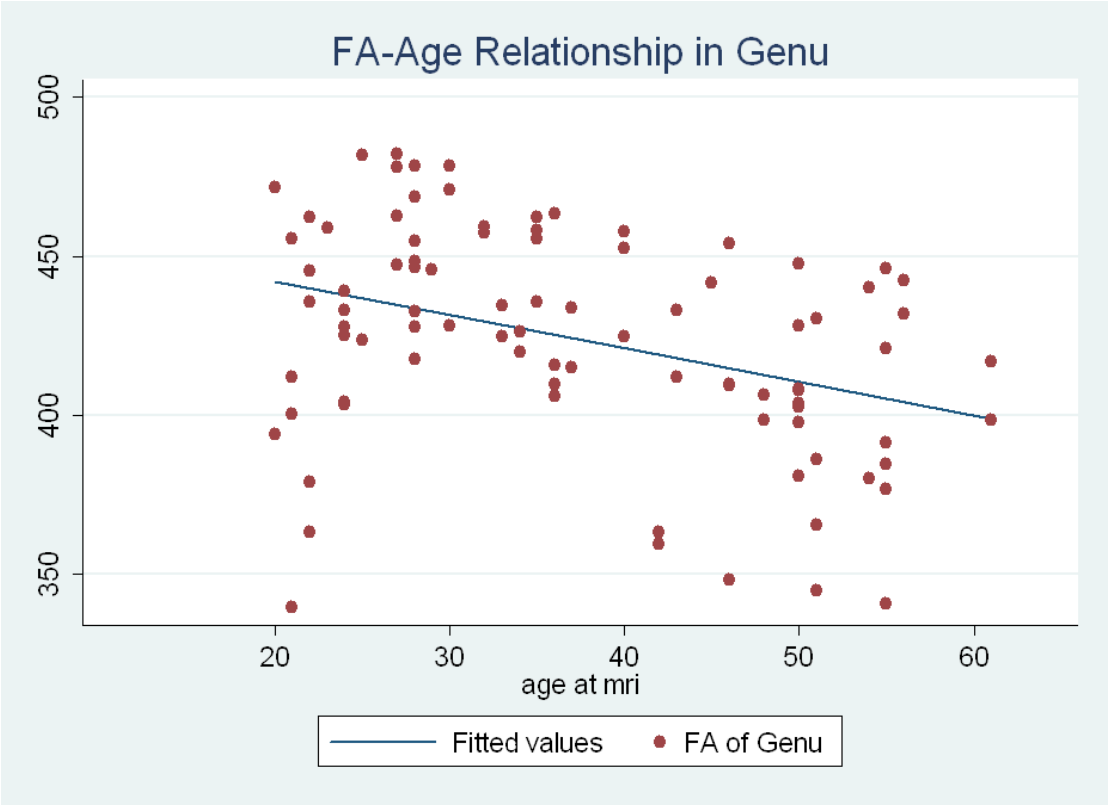

Figure 5.10. Analysis 2a. Relationship Between FA and Age in the Genu

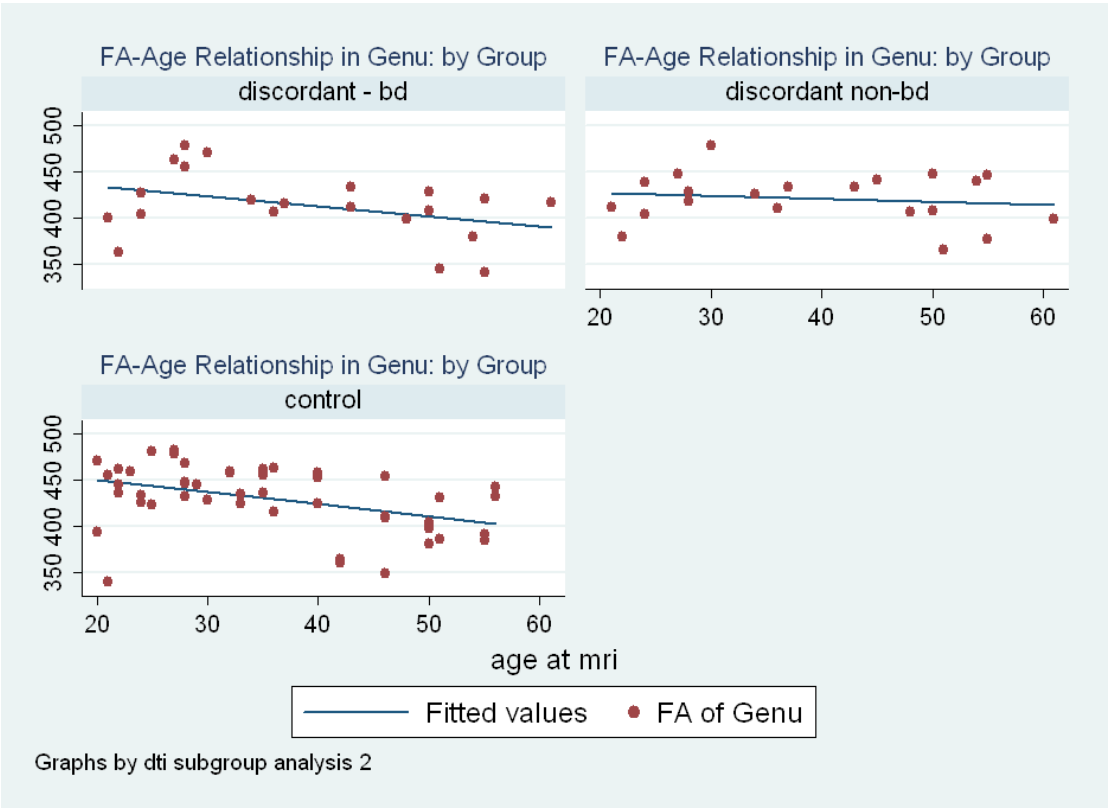

Figure 5.11. Analysis 2a. Relationship Between FA and Age in Genu, by Group

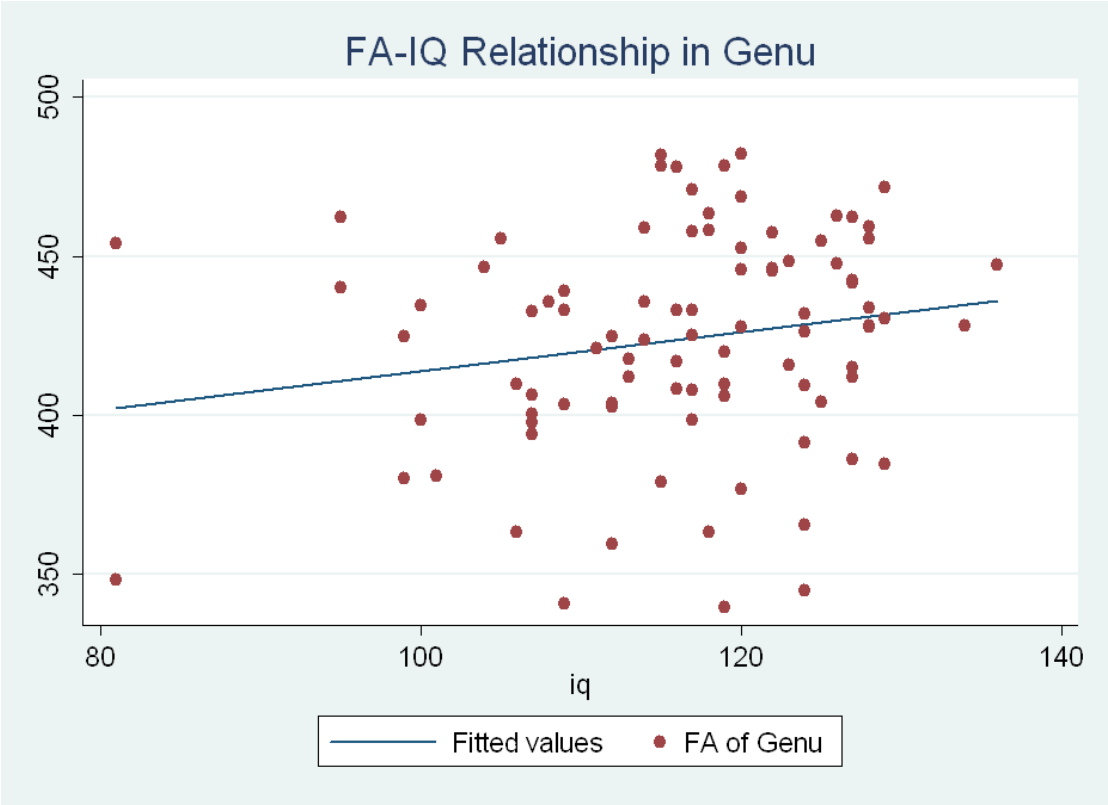

Figure 5.12 Analysis 2a. Relationship between FA and IQ in the Genu

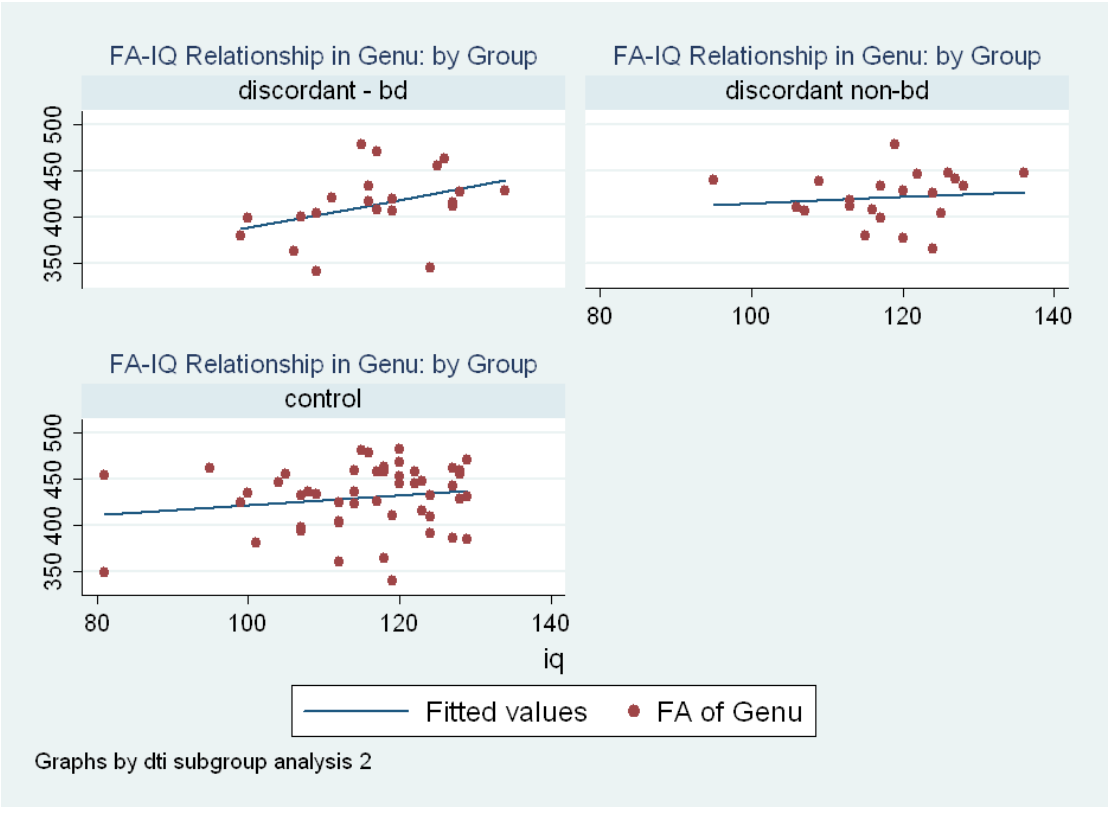

Figure 5.13 Analysis 2a. FA-IQ Relationship in Genu, by Group

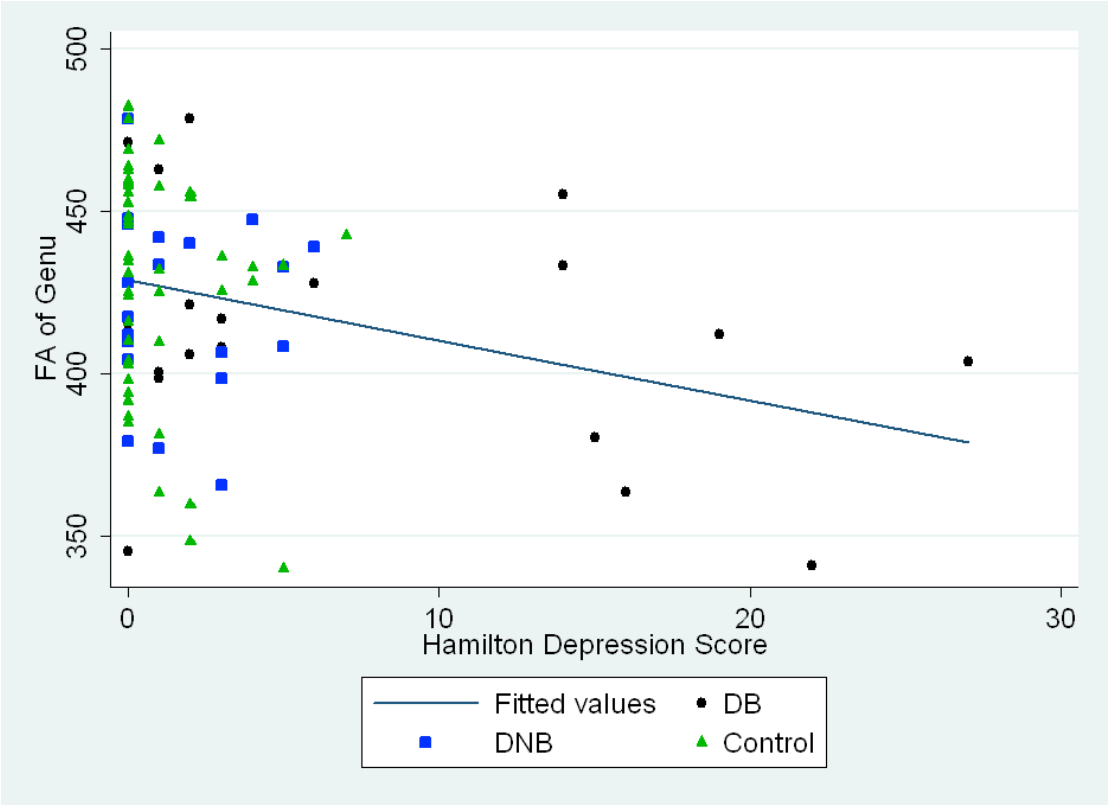

Figure 5.14. Analysis 2a. Relationship between FA and HamD in the Genu

Table 5.10 Analysis 2a. Demographic and Mood Characteristics, with Regression Statistics.

	Group			Test	DB-NDB				DNB-Control			
	Discordant BD (DB)	Discordant Non-BD (DNB)	Control		Sig	C.I.		Coef.	Sig	C.I.		Coef.
						Low	High			Low	High	
Demographics												
N	21	21	49									
Gender (% male)	28.6	28.6	22.4	c2	1.000							
Ethnicity (% white cauc, other)	95.2	95.2	87.8	FE	1.000							
Handedness (% left, right, mixed)	80,15,5	95,5,0	87.8,4.1,8.2	FE	0.342							
PSC (% I, II, III, IV, V, UE)	19,38.1,33.3,4.8,4.8,0	19,38.1,33.3,4.8,4.8,0	16.3,34.7,32.7,6.1,6.1,4.1	FE	1.000							
Age (sd)	39.1 (12.8)	39.2 (12.9)	35.7 (11.2)	reg	0.335	-0.106	0.296	0.095	0.335	-10.785	3.751	-3.517
IQ (sd)	116.7 (9.5)	118.0 (9.1)	115.2 (11.3)	reg	0.349	-1.564	4.231	1.333	0.337	-8.741	3.054	-2.844
Years of education (sd)	16.2 (2.5)	16.6 (2.6)	15.5 (2.6)	reg	0.584	-0.915	1.581	0.333	0.160	-2.606	0.443	-1.082
Diagnois: BD I, BD II	20,1											
Age of onset of mania/hypomania (sd)	26.8 (10.4)											
Years since onset (sd)	12.3 (10.0)											
Mood												
HAM-D (sd)	7.4 (8.6)	1.7 (2.0)	.9 (1.6)	reg	0.007	-9.673	-1.727	-5.700	0.152	-1.813	0.291	-0.761
YOUNG-M (sd)	2.6 (3.6)	1.2 (2.7)	.4 (.9)	reg	0.134	-3.274	0.474	-1.400	0.227	-1.986	0.484	-0.751

Key to Abbreviations: c2: Chi Squared, FE: Fishers Exact, reg: regression, PSC=parental social class (groups I-V, unemployed), HAM-D: Hamilton Depression Scale, Young-M: Young Mania Scale

Table 5.11. Analysis 2a: FA and MD, by Group and Region, With Regression Statistics and Effect Sizes

		Mean FA/MD (s.d.)						DB-DNB Comparison				DNB-Control Comparison				D-Control
		DB		DNB		Control		Sig	C.I.		Coef.	ES	Sig	C.I.		Coef.	ES	ES
FA	**Splenium**	0.482	0.027	0.480	0.036	0.503	0.039	0.764	-18.889	14.070	-2.409	0.08	0.024	3.136	43.448	23.292	-0.62	-0.62
	SLF	0.371	0.039	0.381	0.042	0.394	0.037	0.247	-7.141	26.184	9.521	-0.24	0.236	-9.083	35.971	13.444	-0.34	-0.60
	Genu	0.414	0.037	0.421	0.027	0.430	0.035	0.223	-4.814	19.410	7.298	-0.23	0.293	-7.855	25.453	8.799	-0.28	-0.45
MD ($x10^{-6}$ mm^2/s)	**Splenium**	0.940	0.125	0.937	0.215	0.866	0.100	0.958	-117.014	111.143	-2.935	0.02	0.165	-172.077	30.357	-70.860	0.42	0.65
	SLF	0.746	0.041	0.736	0.033	0.732	0.028	0.271	-27.107	8.044	-9.532	0.26	0.650	-21.793	13.734	-4.030	0.13	0.38
	Genu	0.807	0.065	0.821	0.120	0.775	0.037	0.592	-39.334	67.084	13.875	-0.14	0.095	-99.951	8.226	-45.863	0.52	0.61

Abbreviations: DB – discordant-bipolar, DNB – discordant-non-bipolar, ES - Effect Size (Cohen's D), CC - corpus callosum, SLF - superior longitudinal fasciculus

5.3.3. Analysis 2b: Further Investigation of FA Differences in the Splenium

The difference between bipolar patients and controls from analysis one, combined with the significant DiscNB-control difference from analysis 2 suggest that FA reductions in the splenium might represent an endophenotype of bipolar disorder. To further investigate this difference, the DB and DNB groups were subdivided by zygosity. This resulted in 4 groups: DB-MZ, DB-DZ, DNB-MZ and DNB-DZ

Linear regression was used to investigate potential differences in FA between DB-MZ and DNB-MZ twins as well as between DB-DZ and DNB-DZ twins. There were no significant differences in either comparison. Mean FAs for each group are plotted in Figure 5.15 and shown (with regression statistics) in Table 5.12.

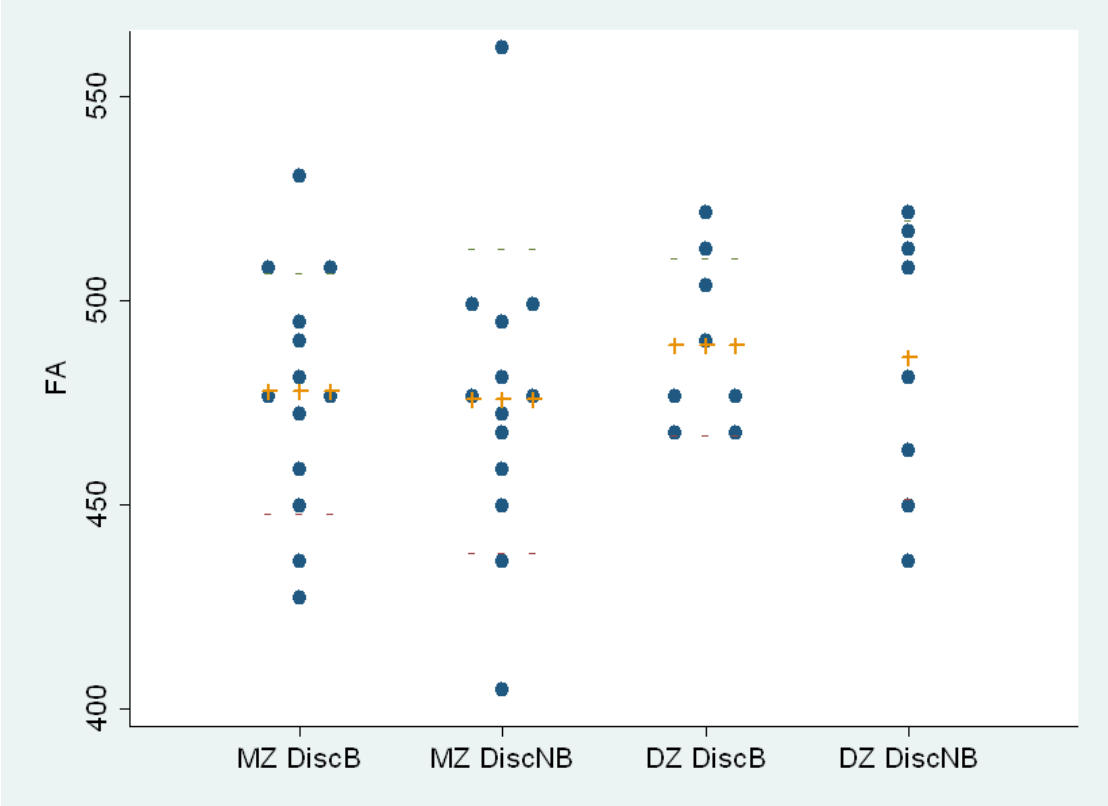

Figure 5.15 Analysis 2b. FA of the Splenium by Group

Table 5.12 Analysis 2b. Mean FA, by Group, With Regression Statistics

	Mean FA (s.d)				MZ DiscB - DiscNB				DZ DiscB - DiscNB			
	MZ		DZ		Sig	C.I.		Coef.	Sig	C.I.		Coef.
	DB	DNB	DB	DNB		low	high			low	high	
Splenium	0.478 (0.029)	0.476 (0.037)	0.489 (0.022)	0.486 (0.034)	0.854	-24.688	20.766	-1.961	0.820	-34.634	28.357	-3.138

5.4. Discussion

In agreement with the primary hypothesis, the study found evidence of reduced FA in patients relative to controls in three regions, which included of the following areas: bilateral genu, bilateral internal capsule, left corpus callosum (CC) body, left anterior splenium, left inferior longitudinal fasciculus (ILF) and left superior longitudinal fasciculus (SLF). These findings are in general agreement with the published literature, which also includes findings of reduced FA in the ILF[220], SLF[135] and CC body[135,139]. The implications of these differences are discussed later.

The second hypothesis was that differences found in patients with bipolar disorder would also be present in the unaffected co-twins of twins with bipolar disorder. Consistent with this hypothesis, in analysis 2a, the study also found evidence that in the splenium/ILF (but not the genu/CC and SLF) region, the DNB group had significantly reduced FA relative to the control group. In this region, FA of the DNB group was indistinguishable from that of the discordant-bipolar group.

While the DNB-control group differences in the genu and SLF regions were not significant, it is worth nothing that, in both these regions, the mean values for the DNB group are intermediate between those of the DB and control groups (see Figure 5.7 and Table 5.10), the effect sizes of these differences being small-medium (using Cohen's[28] criteria). Clearly, such intermediate differences are, by their very nature, difficult to detect - and it is possible that while the study had sufficient power to detect differences between DB and control groups, it did not have enough power to detect smaller differences.

The third hypothesis was that FA differences between discordant MZ twin pairs would be less than FA differences between discordant DZ twin pairs. This was addressed in analysis 2b for the splenium region where an FA difference was detected in both the DB-control and DNB-control comparisons. However, in this analysis, no significant differences were detectable between either discordant MZ or between discordant DZ pairs.

Demographic/Mood Correlations With FA

FA-Age

Analysis 1 revealed a significant inverse relationship between age and FA in all regions, but when decomposed by group, this relationship only held for the genu, where it remained significant for both controls and patients. In analysis 2a, there was also a significant age-FA relationship in the genu, but this was only significant in the control group. It is possible that the DB and DNB groups were simply too small to detect a relationship. The finding of an age-FA inverse relationship is in general agreement with the literature on DTI and aging[221-224].

FA-IQ

Analysis 2a revealed a significant FA-IQ relationship, but only in the DB group. It is only possibly to speculate on this dissociation, but it is plausible that in healthy subjects, the FA of the genu has little effect, but in patients, who demonstrate FA abnormalities, the degree of FA reduction reflects an underlying pathology that also affects IQ.

FA-Depression Scores

In both analyses, significant a significant negative relationship was seen between depression scores and FA. However, on further analysis, it appears that this relationship was driven primarily by the combined group differences in FA and depression scores, rather than a true FA-depression relationship. This illustrates quite nicely the potential problems of trying to 'control' for variables such as mood scores by using them as covariates (as discussed earlier). Here, because the groups differ significantly on their mood scores, the mood scores have in effect become a proxy for group membership and have resulted (most probably) in type I errors.

Interpretation of FA differences.

The following is a discussion of the possible consequences of white matter abnormalities in the regions identified in the present analysis: the superior longitudinal fasciculus, splenium, corpus callosum body, genu and inferior longitudinal fasciculus (as part of the splenium cluster).

Superior Longitudinal Fasciculus

The SLF is a major association pathway linking the temporal and parietal association areas to the frontal lobe. In non-human primates, the SLF has been shown to consist of four components (SLF I,II,II and arcuate fasciculus (AF)) of which SLF II is the major component[225]. Due to the obvious experimental restrictions on anatomical studies in humans, identifying analogous subdivisions in the human brain has historically proved difficult. However, a number of studies have taken advantage of DTI techniques to investigate the human SLF in vivo[226,224,227,228]. One of these studies (Makris et al[224]) specifically investigated whether such subdivisions could be identified in humans using DTI; the authors provided convincing evidence that this was possible and set out where they believed these subdivisions were located. Using the conclusions of Makris et al's study as a guide, the region of reduced FA identified in the present study most likely consists of SLF II and/or arcuate fasciculus. As the anterior-posterior section of the arcuate fasciculus runs parallel to SLF II, it is not possible to make more specific judgements from the present analysis.

SLF II is a major link between the prefrontal cortex and the parietal lobe. The connection is bidirectional, and so in one direction may provide the prefrontal cortex with perceptual information, while in the other direction, may facilitate prefrontal regulation of perceptual attention. The arcuate fasciculus connects caudal superior temporal gyrus to the posterior dorsolateral prefrontal cortex. The superior temporal gyrus is involved in the processing of auditory information and according to Makis et al[224], the AF likely conveys both visuo-spatial and audio-spatial information to the prefrontal cortex

Given the above, we can speculate on the possible consequences of damage or developmental abnormalities in these areas. Alterations in SLF II may result in deficits in both the passage of visuo-spatial information to the prefrontal cortex and the mediation of attention in response to such information. It is also suggested[229] that

such damage could result in deficits with spatial working memory, due to connections with Brodmann area 46 (roughly equivalent to the dorsolateral prefrontal cortex[230]). In the AF, abnormalities might hypothetically result in deficits in spatial processing in both the visual and auditory domains. Both visuo-spatial and verbal deficits have been reported in bipolar disorder and it is tempting therefore to suggest that at least some of the observed deficits may be due to abnormalities in the subcomponents of the SLF. An investigation of the relationship between DTI measures in the SLF and measures of verbal function and visuo-spatial processing might shed some light on this hypothesis.

Corpus Callosum: Splenium and Genu

The corpus callosum (CC) in the largest white matter structure in the brain, consisting of between 200 and 800 million axonal fibres[231]. The main function of the corpus callosum is to control of information transfer between hemispheres[232], and this is clearly reflected in its structure. The fibres of the CC are both heterotopic and homotopic[16] and may serve either excitatory or inhibitory functions. It remains unclear as to whether inhibition or excitation is the CCs primary function[232], although the consensus appears to lean towards an excitatory role (primarily facilitating integration of information rather than facilitating independent function of each hemisphere).

The connections of the corpus callosum are organised in a rostral-caudal manner, facilitating lateral connection of those areas most proximal each section. Based on a dissection study, Witelson[233] suggested that the CC could be divided into 5 sections. A recent DTI study by Hofer and Frahm[234] produced found evidence to confirm this, although the authors proposed some alterations to Witelson's divisions and classifications. Hofer and Frahm divided the CC into (working caudally): Region I: prefrontal; region II: premotor and supplementary motor; region III: motor; region IV: sensory and region V: parietal, temporal, and occipital (Figure 5.16). Using Hofer and Frahm's classifications, the genu/body region identified in the current study

---

[16] In neuroanatomy, a homotopic connection is one that connects an area on one hemisphere to the same area in the contralateral hemisphere. Non-homotopic connections are heterotopic.

corresponds to bilateral section I as well as left region II and III. The splenium region is located in left section V.

The full extent of the specific functional roles played by each section of the CC remains unclear. Much of the evidence that does exist comes from the literature on callosectomy surgery for treatment of epilepsy, as well as the literature on corpus callosum agenesis. Callosectomy involves full or partial surgical lesioning of the CC, typically in order to prevent the spread of abnormal brain activity during epileptic fits. Partial or two stage callosectomies are often carried out in order to minimise functional impairment, and these provides some insight into the differing functions of CC segments. In some cases, the evidence fits with what we know of the primary functionality of the cortical areas subserved by each section. For instance, the splenium connects parietal, temporal and occipital regions, which are commonly associated with spatial, auditory and visual processing (respectively). Thus, it is logical that the CC would be involved in the inter-hemispheric integration of information from these areas – and indeed, sectioning of the splenium (section V) often results in sensory disconnection syndrome (in which intra-hemispheric association of visual and verbal information is interrupted) and alexia[235]. The consequences of prefrontal cortex (genu) lesions are more difficult to understand. Such lesions are reported to result in a transient disconnection syndrome similar to supplementary motor are syndrome and characterised by a variety of symptoms such as non-dominant paresis and mutism[235]. It is possible that it is simply more difficult to interpret the results of prefrontal lesion due the less well defined functions of the prefrontal cortex.

Agenesis of the corpus callosum (AgCC) is the failure of the corpus callosum to develop; it can be either full or partial and occurs in about 1:4000 live births. As in patients with callosectomies, AgCC patients exhibit deficits in interhemispheric transfer (IHT) of information. However, some IHT does occur, presumably via other commissural pathways. Paul et al. report that AcCC patients demonstrate complexity dependent IHT difficulties that suggest the CC is necessary for high information load transfers that are not possible via other pathways[236]. Paul et al also suggest that, as abnormalities of the CC (including AgCC) are often reported in schizophrenia and autism, predisposing genes for these disorders may overlap with those for AgCC[236].

Given the increasing evidence of genetic overlap between bipolar disorder and schizophrenia, it might seem reasonable to extend this hypothesis to bipolar disorder as well. In particular, it is interesting to note that homozygous inactivation of the Disrupted in schizophrenia gene (Disc1) has been observed in a strain of mouse (strain 129) known to exhibit CC abnormalities[237,238]. Furthermore, Disc1 has been found to modulate working memory performance in a subset of this strain[239], which may be relevant considering the reports of working memory deficits in bipolar disorder.

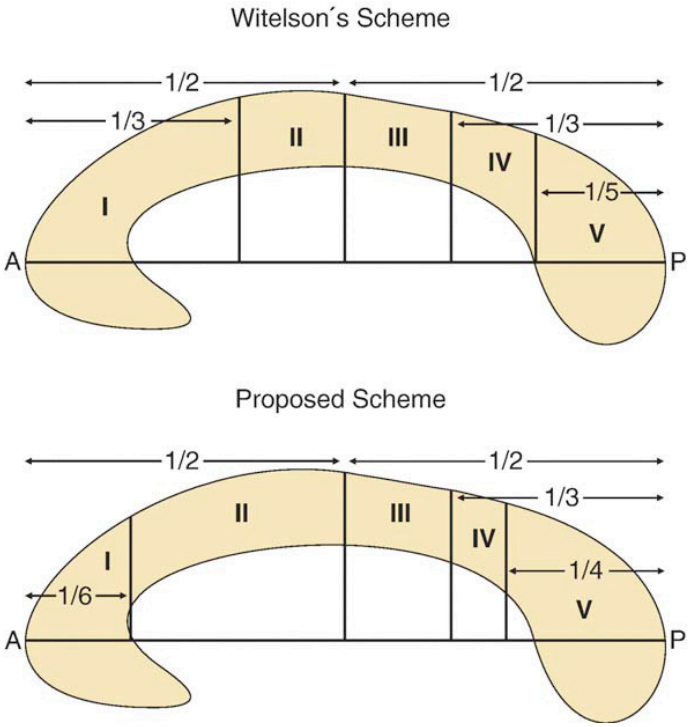

Figure 5.16 Hofer and Frahm's Proposed Subdivisions of the Corpus Callosum

Inferior Longitudinal Fasciculus

The inferior longitudinal fasciculus (ILF) provides a direct connection between occipital and anterior temporal lobes. According to Mandonet et al[240], the ILF runs 'laterally and inferiorly to the lateral wall of the temporal horn… located just laterally and under the optic pathways'. There is actually some controversy as to whether the ILF exists as an independent white matter pathway, but recent evidence suggests that it this is the indeed the case (see Catani et al[241] for a discussion). According to Catani et al[241] the feed-forward (occipital to temporal) role of the ILF may be to facilitate the consolidation of visual memories, while the feedback (temporal to

occipital) role of the ILF may be to provide emotional context for visual processing. The later would provide a plausible mechanism for enabling valence dependent processing of stimuli in the visual cortex.

In the bipolar disorder literature, there are a number of reports of abnormal valance dependent processing of visual stimuli, especially of faces[242]. It is possible that such abnormalities may be due, at least in part, to dysfunction of occipital-temporal connections that allow integration of visual and emotional information. These abnormalities may be state mediated, for example impairments in fear and disgust recognition have been reported in the manic stage[243], while reduced fear recognition with *enhanced* disgust recognition have been reported in the euthymic stage[100]. fMRI studies have found that such abnormalities are associated with over or under activation of the amygdala (relative to controls). Over-activation of the amygdala has been reported in patients observing fearful faces[100], while decreased amygdala activation has been reported in manic patients in response to sad, but not happy, faces[106]. The fact that these abnormalities vary dependent on state means they cannot simply be explained by white matter abnormalities. It is possible however, that connectivity differences due to abnormal white matter may result in processing deficits that increase the emotional lability of patients with bipolar disorder. Unfortunately, the bipolar disorder literature on visual-emotional processing appears to concentrate almost entirely on frontal cortex regulation of temporal structures, at the expense of any investigation of occipital-temporal interaction. Thus, further work is needed to expand this hypothesis.

Possible Weaknesses of the Current Study

Medication

Analysis one used two groups: twin with bipolar disorder and matched control twins. The majority of the subjects in the patient sample were taking at least one psychotropic medication, while none of the control subjects were taking such medication. It is therefore possible that the differences identified were due to medication effects rather than bipolar disorder itself.

By contrast, analysis 2a demonstrates one of the major strengths of family and twin studies; differences detected in unaffected, unmedicated family members cannot be due to medication. In this case, two of the DNB participants were taking psychiatric medication, but this represents less than ten percent of the group. Indeed, when the analysis was repeated without these participants, the results were not significantly changed.

Sample Size

While the sample used in analysis 1 was sufficient to detect significant between-group differences, the samples used in the later analyses (especially 2b), were probably underpowered to the intermediate differences one might expect to find in unaffected relatives. As, at the onset of this study, there was no published research using DTI to investigate bipolar disorder, the magnitude of such effects was unknown. The results of this and other similar studies should help assess the sample sizes required in future studies. It is important to note however, that in the present study, the limiting factor was simply the prevalence and availability of twins with bipolar disorder. In future, it would be sensible to use a combination of twin and family study methodologies, which should make recruitment of large samples more feasible. Fortunately, at the Institute of Psychiatry, a separate team have collected equivalent DTI data from a family study of bipolar disorder and we shall conduct a further analysis with this combined sample.

Sample Characteristics

As noted in the methods chapter, the patients that took part in the current study were recruited nationwide and many were self-referred. As a result, it may be that this sample was less impaired than those from other studies. This may be considered as weakness insofar as it may have reduced our power to detect differences that would

be evident in a more impaired sample. Alternatively, it may be argued that our sample was more reflective of the overall clinical population.

Study Design

The current study was cross-sectional in design and therefore it was not possible to draw firm conclusions about how the detected changes in patients may have evolved. Clearly however, abnormalities seen in the discordant-not-bipolar group cannot be dependent on the expression the disorder itself, but may represent an underlying vulnerability to the disorder.

Mania and Depression Levels

A common criticism of studies of psychiatric disorders is that they do not take into account patients' current mood states. It is often required that researchers attempt to control for such symptoms by using them as covariates. However, this is statistically controversial approach (see 2.11) and so has not been attempted here. It seems unlikely that current mood state would have a significant effect on DTI (as opposed to fMRI) data; although it must be added that there is no published research to validate this assumption.

Other

Both the scanner and the scanning parameters used in this study would now be considered somewhat outdated. While the current study was conducted using a 1.5 Tesla scanner, 3 Tesla scanners are becoming more common; these higher field strength scanners have an inherently higher signal to noise ratio, which can be 'traded in' for higher spatial resolution, shorter scan times, or a combination of both. They suffer, however, from a number of disadvantages, the most serious being increased distortions in EPI based scans such as fMRI and DTI, and increased SAR (specific absorption rate) meaning that power deposition must be carefully managed for subject comfort and safety. Luckily, methodological advances such parallel imaging techniques are overcoming many of the difficulties of higher field strengths, allowing high quality data to be collected in relatively short scan time.

6. Summary and Future Work

6.1. Summary of Main Findings

6.1.1. No fMRI Abnormalities in Bipolar Disorder During A Working Memory Task

The thesis found evidence of behavioural deficits, but not fMRI or connectivity abnormalities during the N-back working memory task. The fMRI finding is in partial contradiction with the published literature, which has generally, but inconstantly reported fMRI abnormalities in bipolar patients performing working memory tasks.

By the standards of the fMRI literature, the study had a large sample size, in both groups – indeed it was the largest sample to date. It was therefore not underpowered to detect differences between patients and controls. Indeed patient-control differences have been detected in considerably smaller samples, using the same task. The analysis method used for the fMRI analysis was clearly able to identify the brain networks involved in performance of the N-Back task and thus should have also been able to detect any group differences that were present. Thus, the absence of group differences is difficult to attribute to an inability of the software to identify differential activation. Although the sample in the present study was large, it was heterogenous with respect to age, gender, age of onset of illness, duration of illness, and type of bipolar disorder (some had no history of psychosis, some were BD-II). Moreover the patient group had been chronically exposed to antipsychotic medication, antidepressants and mood stabilisers, all of which can alter cognitive performance and the BOLD response. Finally, a conservative statistical approach was employed, with a formal Bonferroni-like correction for multiple comparisons. None of the prior studies in the literature have used such a strict statistical approach, but this leaves them very vulnerable to false positive findings.

While deficits of cognitive function are associated with bipolar disorder, they are generally relatively mild in nature, and may be confounded by attendant deficits in attention. Given the nature of the disorder, it is likely that emotional tasks, which tap into the core features of bipolar disorder, may prove more reliable at revealing activation and connectivity differences.

6.1.2. White Matter Abnormalities Are Associated With Bipolar Disorder

The thesis found evidence that bipolar disorder was associated with abnormalities of white matter in areas corresponding to the genu, splenium and body of the corpus callosum, superior longitudinal fasciculus and inferior longitudinal fasciculus. Furthermore, in the splenium of the corpus callosum, abnormalities of white matter were observed in the unaffected co-twins of bipolar twins. Thus the thesis found evidence not only of abnormal white matter in bipolar disorder, but also of a familial predisposition to white matter disorder. Thus, the thesis supports the notion that white mater abnormalities in bipolar disorder represent a *potential* endophenotype for the disorder. However, the thesis was not able to establish the degree to which the observed white matter abnormalities are genetically or environmentally mediated, therefore, such abnormalities remain potential, rather than confirmed endophenotypes.

Failure to establish the extent of the genetic influence on the observed white matter deficits was primarily due the sample size, in the final analysis, which was too small to perform any genetic analysis. Recruiting large numbers of twin pairs discordant for bipolar disorder is a lengthy and difficult task, especially is countries without comprehensive and readily available psychiatric registers. Although further work is necessary, if confirmed as endophenotypes for bipolar disorder, white matter deficits in the areas identified may be used as quantitative phenotypes for the identification of susceptibility genes for bipolar disorder.

6.2. Future Work

In order to fully elucidate the findings from this thesis, it will be necessary to carry out considerable future work. A range of other data has been collected from the same subjects investigated in this thesis, including two other fMRI paradigms (emotional faces and verbal fluency), a neuropsychological battery, structural imaging data and genetic samples. One of the main strengths of the Maudsley Bipolar Twin Study lies in the multiple different investigations that have been conducted in the same sample. DTI data by itself is useful and interesting, but by combined it with structural, fMRI and neuropsychological data, it should be possible to provide more nuanced explanations of the findings from each modality. Possible future work is therefore discussed below.

6.2.1. Integration of DTI, fMRI and Structural Imaging Findings

The DTI analysis in this thesis was performed after the N-Back fMRI and N-Back connectivity analyses. However, given that the DTI and fMRI data was gathered in the same sample, it is logical to use the results of one analysis to inform the other. In particular, the finding of reduced FA in both patients with bipolar disorder and their unaffected cotwins has implications for connectivity. White matter pathways are, of course, the primary means by which information is passed from one part of the brain to another part of the brain. Thus any abnormality of white matter will presumably result in a corresponding abnormality of connectivity (although it is possible that the brain may adapt to white matter abnormalities by mechanisms such as neurotransmitter receptor upregulation).

The reduced FA observed in the superior and inferior longitudinal fasciculi may plausibly result in altered connectivity between frontal, parietal and temporal cortices. Reduced FA in the splenium and genu of the corpus callosum, is perhaps more likely to result in altered connectivity between hemispheres. Although two connectivity analyses have been performed as part of this thesis, without activation differences or DTI findings to guide them, they were necessarily exploratory. A model driven analysis, with specific hypotheses, may have more power to detect subtle differences in connectivity. I therefore propose to use Dynamic Causal Modelling[244] (DCM) to test whether the white matter abnormalities observed in the corpus callosum and SLF result in specific deficits in connectivity.

The DTI findings reported in this thesis also have implications for the analysis of the two other fMRI paradigms used in the Maudsley Bipolar Twin Study. In particular, I believe (as noted earlier) that, given the nature of bipolar disorder, it is more likely that emotional will result in abnormal brain activation than pure cognitive tasks. Thus, it is possible that combining DTI and the data from the emotional faces task may prove more fruitful than may be the case with the working memory task analysed in this thesis.

Finally, SPGR structural brain images have been collected for all of our subjects. Given the finding of altered white matter in the DTI analysis, it is likely that

abnormalities will also be present in the structural imaging data. I anticipate taking a dual methodology approach to analysis of the structural imaging data. Firstly, a region of interest study will be conducted, using both the DTI findings and the results of the recent meta-analysis by Kempton et al[59] as a guide to the most likely regions of difference. Secondly, a voxel based analysis will be conducted, to ensure that areas of difference.

6.2.2. Integration of Findings from Twin and Family Studies

Sample size is a key issue in imaging and neuropsychological studies in bipolar disorder. Meta-analyses have reported that many of the reported deficits that may be present in bipolar disorder may be relatively subtle[27,59,25,26] - and therefore their reliable detection requires considerably larger same sizes than are the norm. There are several affiliated studies at the Institute of Psychiatry that have collected DTI data using the same protocol and scanner – and which have also collected compatible neuropsychological data. These studies include the Maudsley Family Study of Bipolar Disorder and the Maudsley Schizophrenia Twin Study. A clear next step is to combine the data from all of these studies. Not only will this increase the statistical power of any analysis, but it will also allow more sophisticated analysis such as structural equation modelling[245], a technique that allows one to model shared variance due to genetic and environmental influences and thus to evaluate potential endophenotypes.

6.2.3. Integration of DTI and Neuropsychological Findings

Elucidation of possible relationships between neuropsychological variables and imaging data has the potential to greatly aid interpretation of imaging findings. To date however, no studies have been published investigating the relationships between DTI and neuropsychological variables in bipolar disorder. In schizophrenia, there are three relevant papers. Two of the studies, by Nestor et al[246,247] investigated correlations between uncinate fasciculus (UF) and cingulate bundle FA and neuropsychological variables, in both patients and controls. The studies found that, in patients with schizophrenia, but not controls, there was (i) a significant relationship between (i) FA of uncinate fasciculus and declarative memory and (ii) a significant relationship between FA of cingulate bundle and verbal function, performance IQ and working memory. The findings of Szeszko et al[248] confirm these data, reporting

significant relationships between FA of the UF and verbal memory, logical memory, and verbal fluency as well as negative symptoms, alogia and affective flattening.

As part of the Maudsley Bipolar Study, a battery of neuropsychological tests was performed with each volunteer. This provides an opportunity to investigate the relationship between white matter as measured by DTI and neuropsychological variables. Although the current thesis did not specifically investigate the relationship between neuropsychological variables and DTI measures. Nevertheless, for the one neuropsychological variable that was investigated, IQ, a significant positive relationship was found with the FA of the genu of the corpus callosum (however, this relationship would not have survived a Bonferroni correction for multiple testing in the analysis of demographic variables).

6.2.4. Investigation of Specific Genetic Influences on White Matter and fMRI

A number of the genes so far identified with risk for bipolar disorder normally play an important role in the development of white matter. These include the genes coding for brain derived neurotrophic factor (BDNF)[249] and neuregulin-1 (NRG-1). NRG-1 and its receptor Erb4 in particular are extremely important in brain development and neural plasticity[250,251]. McIntosh et al[252], using both DTI and structural imaging have shown that, in normal controls, variation in the NRG-1 gene is associated with white matter changes in the anterior limb of the internal capsule. Thus it is of great interest to investigate whether a similar effect may be present in the patient population – or indeed whether these genes have a differential effect in patients, relative to controls. Realistically, to have power to detect gene-brain interactions, such a study would require greater numbers than are available in the Maudsley Bipolar Twin Study sample, so this analysis would ideally be conducted in a combined sample as described earlier.

6.2.5. Investigation of Possible Singleton-Twin Structural/White Matter Differences

In order to combine data from twin and singleton studies (as discussed above), it is important to know whether there are twin-singleton differences across the variables of interest. It is currently unknown whether twins differ from singletons in terms of brain morphology or white matter integrity. However, given the increased rate of obstetric complications in twins, combined with the shared environment in utero, it is

quite possible that twin birth may give rise to altered brain development. Studies of IQ differences between twins and singletons appear to support this hypothesis; a recent meta-analysis by Voracek and Haubner[176] reports a mild (4.2 IQ points), but statistically significant reduction of IQ in twins compared to singletons. Although this difference is only 1/3 of a standard deviation, it presumably represents an abnormality of brain development. I therefore propose to use DTI scans, from the Maudsley Bipolar Twin Study and other studies conducted at the Institute of Psychiatry in order to investigate the presence and magnitude of putative twin-singleton brain differences.

6.2.6. How The Study Would be Run Today.

Taking into account developments since this thesis was started, and with the benefit of hindsight; if starting the study today, there are a number of key changes I would make the way the study was designed and run. These are detailed below.

Sample

The most significant change involves the nature of the sample. As noted earlier in the thesis, twin samples in psychiatric populations are, due to their scarcity, very difficult to recruit. The advantages of twin samples in terms of being able to partition genetic and environmental influences on outcome variables are traditionally considered enough to make recruitment of twin samples worthwhile. However, recent developments in study design mean that family studies may now be able to offer the same advantages as twin studies. This is primarily due to the development of genetic liability scales (such as that designed by Professor Sham and used by McIntosh et al[74] and McDonald et al[73]) as well as modelling techniques that allow for partitioning of genetic and environmental factors in complex extended pedigrees. It has also been proposed the power of twin studies can be significantly enhanced by supplementing twin pairs with other family members[253]. These techniques also offer the ability to investigate factors such as nonrandom mating and genotype X environment covariation. Thus, if beginning the study now, I would recruit both extended family pedigrees and twin pairs in order to overcome the difficulties associated with twin recruitment and to increase the power available for analysis.

Imaging

As mentioned earlier, the scanner and parameters used in the current study might now be considered sub optimal, particularly with regard to the fMRI data collection. Newer scanners and techniques promise a number of advances including higher field strengths, better signal to noise ratio, shorter scan times and higher resolutions[254]. There will always, however, be compromises and trade offs in magnetic resonance imaging, whereby one parameter may be optimised at the expense of another (e.g. resolution vs. scan time). A recent development that is particularly exciting is the ability to obtain DTI images that allow one to address the problem of crossing fibers in[255] (this is an issue whereby crossing fibers both affect observed levels of fractional anisotropy (FA) and prevent compromise tract tracing). This development should help researchers to further explore what lies behind the observed differences of white

matter in disorders such as bipolar disorder and schizophrenia. Beyond this, there have also been recent developments in whole brain analysis of DTI data, such as Tract Based Skeletal Statistics (TBSS) that promise to provide more accurate registration than was available with previous techniques[142]. Clearly, if starting the study now, I would be keen to take advantage of as may of these recent developments as possible.

Diagnostic Tools and Clinical Assessment

I would no longer use the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) to assess psychiatric diagnosis of volunteers, I would now use a tool such as the Structured Clinical Interview for DSM-IV (SCID-I). The SCAN interview, while a useful tool, proved to be rather unwieldy, both because it took a long time to administer and because its computerised diagnostic algorithms were subject to glitches that would sometimes produce spurious diagnoses. While the algorithms are intended to produce a less administrator biased and more reliable diagnosis, they do not currently live up to this promise. After many discussions with colleagues, I am now of the opinion, for a study such as the current, other tools such as the SCID-I offers a better balance of convenience, accuracy and rigour. Finally, if running the study now, I would incorporate a scale assessing the lifetime severity of a patient's disorder, such as the he Bipolar Affective Disorder Dimension Scale (BADDS)[256]. I believe that the use of a formal severity scale would help to better describe the clinical characteristics of the sample. Furthermore, it would help in any attempt to correlate illness severity with cognitive or neuroimaging abnormalities.

6.3. Concluding Remarks

Two of the studies in my thesis failed to find any evidence of the putative neurobiological differences associated with bipolar disorder. The remaining study, however, found evidence not only of neurobiological differences in patients, but also in their unaffected relatives, suggesting that white matter deficits may represent an endophenotype of bipolar disorder. These findings are, of course based on statistical probability and represent not truth, but evidence. The positive findings may be false positives and the negative findings may be false negatives. Only time can tell.

For me, the frustrations of psychiatric research are often immense. At times, it is impossible to know what to believe, both from the literature and, I have learnt, in one's own research. In the literature, it is virtually impossible to quantify the impact of factors such as pressure to publish and positive publication bias. It is clear that these factors distort the literature, but rarely clear how much. From my perspective as an individual researcher, while I have endeavoured to be completely honest in the writing of this PhD, I, like all researchers am subject to bias and error. It is perhaps inevitable that as part of the process of a PhD, one becomes increasingly aware of the problems inherent in the literature, as well as the limitations of one's own methods. As the American writer and philosopher Will Durant stated, "Inquiry is fatal to certainty"; this appears true, but it can be hard to live with. In psychiatric research, when the very concepts upon which research is founded are shaky, it appears particularly hard – although this is probably an artefact of perspective and familiarity. Given the understandable human need for a sense of purpose and mastery, when faced with these issues it can be difficult to maintain enthusiasm and indeed, to avoid despair.

The history of science tells us that there are no inexorable truths. While I have long known this, truly understanding it has been a more difficult process. However, coming to terms with this concept and its implications, which for a scientist must also come with humility, has been essential. Essential to overcome cynicism and regain the sense of curiosity that makes science worthwhile and scientists productive. In truth, while the contribution of this thesis to human knowledge is infinitesimal, for me, is has been the process that has most valuable. From frustration, fruition.

7. References

1. American Psychiatric Association *Diagnostic and statistical manual of mental disorders : DSM-IV-TR*. **4th ed., text revision**, (American Psychiatric Association: Washington, DC, 2000).

2. Weissman, M. et al. Cross-national epidemiology of major depression and bipolar disorder. *JAMA* **276**, 293-299 (1996).

3. Kennedy, N. et al. Gender differences in incidence and age at onset of mania and bipolar disorder over a 35-year period in Camberwell, England. *Am J Psychiatry* **162**, 257-262 (2005).

4. Kennedy, N. et al. Incidence and distribution of first-episode mania by age: results from a 35-year study. *Psychol Med* **35**, 855-863 (2005).

5. Cardno, A. et al. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* **56**, 162-168 (1999).

6. Muller-Oerlinghausen, B., Berghofer, A. & Bauer, M. Bipolar disorder. *Lancet* **359**, 241-247 (2002).

7. Wyatt, R. & Henter, I. An economic evaluation of manic-depressive illness--1991. *Soc Psychiatry Psychiatr Epidemiol* **30**, 213-219 (1995).

8. Murray, C., Lopez, A., Harvard School of Public, H., World Health, O. & World, B. *The global burden of disease : a comprehensive assessment of mortality*. (Harvard School of Public Health on behalf of the World: [Boston, Mass.], 1996).

9. Cardno, A.G., Rijsdijk, F.V., Sham, P.C., Murray, R.M. & McGuffin, P. A twin study of genetic relationships between psychotic symptoms. *Am J Psychiatry* **159**, 539-45 (2002).

10. Berrettini, W.H. Are schizophrenic and bipolar disorders related? A review of family and molecular studies. *Biol Psychiatry* **48**, 531-8 (2000).

11. Lewis, C.M. et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* **73**, 34-48 (2003).

12. Boks, M.P., Leask, S., Vermunt, J.K. & Kahn, R.S. The structure of psychosis revisited: The role of mood symptoms. *Schizophrenia Research* **93**, 178-185 (2007).

13. Keri, S. & Janka, Z. Critical evaluation of cognitive dysfunctions as endophenotypes of schizophrenia. *Acta Psychiatr Scand* **110**, 83-91 (2004).

14. Gottesman, I.I. & Gould, T.D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* **160**, 636-645 (2003).

15. Flint, J. & Munafo, M.R. The endophenotype concept in psychiatric genetics. *Psychol Med* **37**, 163-180 (2007).

16. Gschwandtner, U. et al. Neuropsychological and neurophysiological findings in individuals suspected to be at risk for schizophrenia: preliminary results from the Basel early detection of psychosis study - Fruherkennung von Psychosen (FEPSY). *Acta Psychiatr Scand* **108**, 152-5 (2003).

17. Niemi, L.T., Suvisaari, J.M., Tuulio-Henriksson, A. & Lonnqvist, J.K. Childhood developmental abnormalities in schizophrenia: evidence from high-risk studies. *Schizophr Res* **60**, 239-58 (2003).

18. Liu, S.K. et al. Deficits in sustained attention in schizophrenia and affective disorders: stable versus state-dependent markers. *Am J Psychiatry* **159**, 975-82 (2002).

19. Lewis, R. Should cognitive deficit be a diagnostic criterion for schizophrenia? *J Psychiatry Neurosci* **29**, 102-13 (2004).

20. Nuechterlein, K.H. & Dawson, M.E. A heuristic vulnerability/stress model of schizophrenic episodes. *Schizophr Bull* **10**, 300-12 (1984).

21. Cornblatt, B.A. & Keilp, J.G. Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophr Bull* **20**, 31-46 (1994).

22. Kremen, W.S. et al. Neuropsychological risk indicators for schizophrenia: a review of family studies. *Schizophr Bull* **20**, 103-19 (1994).

23. Davalos, D.B., Compagnon, N., Heinlein, S. & Ross, R.G. Neuropsychological deficits in children associated with increased familial risk for schizophrenia. *Schizophr Res* **67**, 123-30 (2004).

24. Keating, M. et al. Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey ras-1 gene. *Science* **252**, 704-706 (1991).

25. Torres, I.J., Boudreau, V.G. & Yatham, L.N. Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. *Acta Psychiatrica Scandinavica* **116**, 17-26 (2007).

26. Robinson, L.J. et al. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of Affective Disorders* **93**, 105-115 (2006).

27. Arts, B., Jabben, N., Krabbendam, L. & van Os, J. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychol*

Med **38**, 771-785 (2008).

28. Cohen, J. *Statistical power analysis for the behavioral sciences.* **2nd ed.**, (Erlbaum: Hillsdale, NJ, 1988).

29. Balanza-Martinez, V. et al. Neurocognitive endophenotypes (Endophenocognitypes) from studies of relatives of bipolar disorder subjects: A systematic review. *Neurosci Biobehav Rev* **32**, 1426-1438 (2008).

30. Senn, S. Overstating the evidence - double counting in meta-analysis and related problems. *BMC Medical Research Methodology* **9**, 10 (2009).

31. Kravariti, E., Kane, F. & Murray, R.M. Neurocognitive Endophenotypes for Bipolar Disorder: Evidence from Case-Control, Family and Twin Studies. *The Handbook of Neuropsychiatric Biomarkers, Endophenotypes and Genes* 195-210 (2009).

32. Clark, L., Kempton, M.J., Scarnà, A., Grasby, P.M. & Goodwin, G.M. Sustained attention-deficit confirmed in euthymic bipolar disorder but not in first-degree relatives of bipolar patients or euthymic unipolar depression. *Biol Psychiatry* **57**, 183-187 (2005).

33. Glahn, D.C., Bearden, C.E., Niendam, T.A. & Escamilla, M.A. The feasibility of neuropsychological endophenotypes in the search for genes associated with bipolar affective disorder. *Bipolar Disorders* **6**, 171-182 (2004).

34. Robinson, L.J. & Ferrier, I. Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence. *Bipolar Disord* **8**, 103-116 (2006).

35. Ando, J., Ono, Y. & Wright, M. Genetic Structure of Spatial and Verbal Working Memory. *Behavior Genetics* **31**, 615-624 (2001).

36. Bouchard, T.J. Genetic and environmental influences on adult intelligence and special mental abilities. *Hum. Biol* **70**, 257-279 (1998).

37. Finkel, D., Pedersen, N.L., McGue, M. & McClearn, G.E. Heritability of cognitive abilities in adult twins: comparison of Minnesota and Swedish data. *Behav. Genet* **25**, 421-431 (1995).

38. Malhi, G.S. et al. Neuropsychological deficits and functional impairment in bipolar depression, hypomania and euthymia. *Bipolar Disord* **9**, 114-125 (2007).

39. Trivedi, J. et al. Cognitive functions in euthymic state of bipolar disorder: an Indian study. *Cognit Neuropsychiatry* **13**, 135-147 (2008).

40. Breeze, J.L., Hesdorffer, D.C., Hong, X., Frazier, J.A. & Renshaw, P.F. Clinical

significance of brain white matter hyperintensities in young adults with psychiatric illness. *Harv Rev Psychiatry* **11**, 269-283 (2003).

41. Silverstone, T., McPherson, H., Li, Q. & Doyle, T. Deep white matter hyperintensities in patients with bipolar depression, unipolar depression and age-matched control subjects. *Bipolar Disord* **5**, 53-57 (2003).

42. Lyoo, I.K., Lee, H.K., Jung, J.H., Noam, G.G. & Renshaw, P.F. White matter hyperintensities on magnetic resonance imaging of the brain in children with psychiatric disorders. *Compr Psychiatry* **43**, 361-368 (2002).

43. Pillai, J.J. et al. Increased presence of white matter hyperintensities in adolescent patients with bipolar disorder. *Psychiatry Res* **114**, 51-56 (2002).

44. de Asis, J.M. et al. Frontal signal hyperintensities in mania in old age. *Am J Geriatr Psychiatry* **14**, 598-604 (2006).

45. Altshuler, L. et al. T2 hyperintensities in bipolar disorder: magnetic resonance imaging comparison and literature meta-analysis. *Am J Psychiatry* **152**, 1139-1144 (1995).

46. Stoll, A., Renshaw, P., Yurgelun-Todd, D. & Cohen, B. Neuroimaging in bipolar disorder: what have we learned? *Biol Psychiatry* **48**, 505-517 (2000).

47. Pompili, M. et al. White matter hyperintensities and their associations with suicidality in patients with major affective disorders. *Eur Arch Psychiatry Clin Neurosci* **257**, 494-499 (2007).

48. Moore, P. et al. Cerebral white matter lesions in bipolar affective disorder: relationship to outcome. *Br J Psychiatry* **178**, 172-176 (2001).

49. Woods, B., Yurgelun-Todd, D., Mikulis, D. & Pillay, S. Age-related MRI abnormalities in bipolar illness: a clinical study. *Biol Psychiatry* **38**, 846-847 (1995).

50. Sassi, R.B. et al. White matter hyperintensities in bipolar and unipolar patients with relatively mild-to-moderate illness severity. *J Affect Disord* **77**, 237-245 (2003).

51. Ahn, K.H. et al. White matter hyperintensities in subjects with bipolar disorder. *Psychiatry Clin Neurosci* **58**, 516-521 (2004).

52. Strakowski, S. et al. Structural brain abnormalities in first-episode mania. *Biol Psychiatry* **33**, 602-609 (1993).

53. Brambilla, P. et al. Differential effects of age on brain gray matter in bipolar patients and healthy individuals. *Neuropsychobiology* **43**, 242-247 (2001).

54. Davis, K.A., Kwon, A., Cardenas, V.A. & Deicken, R.F. Decreased cortical gray and cerebral white matter in male patients with familial bipolar I disorder. *J Affect Disord* **82**, 475-485 (2004).

55. Hauser, P. et al. MRI-based measurements of temporal lobe and ventricular structures in patients with bipolar I and bipolar II disorders. *J Affect Disord* **60**, 25-32 (2000).

56. Strakowski, S.M. et al. Ventricular and periventricular structural volumes in first- versus multiple-episode bipolar disorder. *Am J Psychiatry* **159**, 1841-1847 (2002).

57. Strasser, H.C. et al. Hippocampal and ventricular volumes in psychotic and nonpsychotic bipolar patients compared with schizophrenia patients and community control subjects: a pilot study. *Biol Psychiatry* **57**, 633-639 (2005).

58. Konarski, J.Z. et al. Volumetric neuroimaging investigations in mood disorders: bipolar disorder versus major depressive disorder. *Bipolar Disorders* **10**, 1-37 (2008).

59. Kempton, M.J., Geddes, J.R., Ettinger, U., Williams, S.C.R. & Grasby, P.M. Meta-analysis, Database, and Meta-regression of 98 Structural Imaging Studies in Bipolar Disorder. *Arch Gen Psychiatry* **65**, 1017-1032 (2008).

60. Lochhead, R.A., Parsey, R.V., Oquendo, M.A. & Mann, J.J. Regional brain gray matter volume differences in patients with bipolar disorder as assessed by optimized voxel-based morphometry. *Biological Psychiatry* **55**, 1154-1162 (2004).

61. Wilke, M., Kowatch, R.A., DelBello, M.P., Mills, N.P. & Holland, S.K. Voxel-based morphometry in adolescents with bipolar disorder: first results. *Psychiatry Research: Neuroimaging* **131**, 57-69 (2004).

62. Adler, C.M., Levine, A.D., DelBello, M.P. & Strakowski, S.M. Changes in Gray Matter Volume in Patients with Bipolar Disorder. *Biological Psychiatry* **58**, 151-157 (2005).

63. Adler, C.M. et al. Voxel-Based Study of Structural Changes in First-Episode Patients with Bipolar Disorder. *Biological Psychiatry* **61**, 776-781 (2007).

64. Chen, X., Wen, W., Malhi, G.S., Ivanovski, B. & Sachdev, P.S. Regional gray matter changes in bipolar disorder: a voxel-based morphometric study. *Aust N Z J Psychiatry* **41**, 327-336 (2007).

65. Ladouceur, C.D. et al. Subcortical gray matter volume abnormalities in healthy bipolar offspring: potential neuroanatomical risk marker for bipolar disorder? *J*

Am Acad Child Adolesc Psychiatry **47**, 532-539 (2008).

66. Nugent, A.C. et al. Cortical abnormalities in bipolar disorder investigated with MRI and voxel-based morphometry. *NeuroImage* **30**, 485-497 (2006).

67. Farrow, T.F., Whitford, T.J., Williams, L.M., Gomes, L. & Harris, A.W. Diagnosis-Related Regional Gray Matter Loss Over Two Years in First Episode Schizophrenia and Bipolar Disorder. *Biological Psychiatry* **58**, 713-723 (2005).

68. Almeida, J.R. et al. Reduced gray matter volume in ventral prefrontal cortex but not amygdala in bipolar disorder: Significant effects of gender and trait anxiety. *Psychiatry Research: Neuroimaging* **171**, 54-68 (2009).

69. Lyoo, I.K. et al. Frontal lobe gray matter density decreases in bipolar I disorder. *Biological Psychiatry* **55**, 648-651 (2004).

70. McIntosh, A.M. et al. Voxel-based morphometry of patients with schizophrenia or bipolar disorder and their unaffected relatives. *Biological Psychiatry* **56**, 544-552 (2004).

71. Bruno, S.D., Barker, G.J., Cercignani, M., Symms, M. & Ron, M.A. A study of bipolar disorder using magnetization transfer imaging and voxel-based morphometry. *Brain* **127**, 2433-2440 (2004).

72. McIntosh, A.M. et al. White Matter Density in Patients with Schizophrenia, Bipolar Disorder and Their Unaffected Relatives. *Biological Psychiatry* **58**, 254-257 (2005).

73. McDonald, C. et al. Association of genetic risks for schizophrenia and bipolar disorder with specific and generic brain structural endophenotypes. *Arch Gen Psychiatry* **61**, 974-84 (2004).

74. McIntosh, A.M. et al. Genetic liability to schizophrenia or bipolar disorder and its relationship to brain structure. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **141B**, 76-83 (2006).

75. Yatham, L.N. et al. A magnetic resonance imaging study of mood stabilizer- and neuroleptic-naïve first-episode mania. *Bipolar Disorders* **9**, 693-697 (2007).

76. Scherk, H. et al. No change to grey and white matter volumes in bipolar I disorder patients. *European Archives of Psychiatry and Clinical Neuroscience* **258**, 345-349 (2008).

77. Uranova, N.A., Vostrikov, V.M., Orlovskaya, D.D. & Rachmanova, V.I. Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. *Schizophrenia*

Research **67**, 269-275 (2004).

78. Uranova, N. et al. Electron microscopy of oligodendroglia in severe mental illness. *Brain Research Bulletin* **55**, 597-610 (2001).

79. Ongur, D., Drevets, W.C. & Price, J.L. Glial reduction in the subgenual prefrontal cortex in mood disorders. *Proc Natl Acad Sci U S A.* **95**, 13290–13295 (1998).

80. Johnston-Wilson, N. et al. Disease-specific alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder. The Stanley Neuropathology Consortium. *Mol Psychiatry* **5**, 142-149 (2000).

81. Regenold, W.T. et al. Myelin staining of deep white matter in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and unipolar major depression. *Psychiatry Research* **151**, 179-188 (2007).

82. Tkachev, D. et al. Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *The Lancet* **362**, 798-805 (2003).

83. Cotter, D., Hudson, L. & Landau, S. Evidence for orbitofrontal pathology in bipolar disorder and major depression, but not in schizophrenia. *Bipolar Disord* **7**, 358-369 (2005).

84. Rajkowska, G., Halaris, A. & Selemon, L. Reductions in neuronal and glial density characterize the dorsolateral prefrontal cortex in bipolar disorder. *Biol Psychiatry* **49**, 741-752 (2001).

85. Chana, G., Landau, S., Beasley, C., Everall, I.P. & Cotter, D. Two-dimensional assessment of cytoarchitecture in the anterior cingulate cortex in major depressive disorder, bipolar disorder, and schizophrenia: evidence for decreased neuronal somal size and increased neuronal density. *Biological Psychiatry* **53**, 1086-1098 (2003).

86. Beasley, C.L. et al. Evidence for altered neuronal organisation within the planum temporale in major psychiatric disorders. *Schizophrenia Research* **73**, 69-78 (2005).

87. Bouras, C., Kovari, E., Hof, P., Riederer, B. & Giannakopoulos, P. Anterior cingulate cortex pathology in schizophrenia and bipolar disorder. *Acta Neuropathol* **102**, 373-379 (2001).

88. Haldane, M.[. & Frangou, S.[. Functional neuroimaging studies in mood disorders. *Acta Neuropsychiatrica* **18**, 88-99 (2006).

89. Baxter, L.R. et al. Cerebral metabolic rates for glucose in mood disorders. Studies with positron emission tomography and fluorodeoxyglucose F 18. *Arch*

Gen Psychiatry **42**, 441-447 (1985).

90. Baxter, L. et al. Reduction of prefrontal cortex glucose metabolism common to three types of depression. *Arch Gen Psychiatry* **46**, 243-250 (1989).

91. Dunn, R.T. et al. Principal components of the beck depression inventory and regional cerebral metabolism in unipolar and bipolar depression. *Biological Psychiatry* **51**, 387-399 (2002).

92. Drevets, W.C. et al. Glucose metabolism in the amygdala in depression: Relationship to diagnostic subtype and plasma cortisol levels. *Pharmacology Biochemistry and Behavior* **71**, 431-447 (2002).

93. Ito, H. et al. Hypoperfusion in the Limbic System and Prefrontal Cortex in Depression: SPECT with Anatomic Standardization Technique. *J Nucl Med* **37**, 410-414 (1996).

94. Tutus, A. et al. Changes in regional cerebral blood flow demonstrated by single photon emission computed tomography in depressive disorders: comparison of unipolar vs. bipolar subtypes. *Psychiatry Res* **83**, 169-177 (1998).

95. Rubin E.[1] et al. Regional cerebral blood flow in mood disorders: IV. Comparison of mania and depression. *Psychiatry Research: Neuroimaging* **61**, 1-10 (1995).

96. Migliorelli, R. et al. SPECT findings in patients with primary mania. *J Neuropsychiatry Clin Neurosci* **5**, 379-383 (1993).

97. O'Connell, R. et al. Single-photon emission computed tomography of the brain in acute mania and schizophrenia. *J Neuroimaging* **5**, 101-104 (1995).

98. Blumberg, H.P. et al. Increased anterior cingulate and caudate activity in bipolar mania. *Biological Psychiatry* **48**, 1045-1052 (2000).

99. Goodwin, G. et al. Uptake of 99mTc-exametazime shown by single photon emission computed tomography before and after lithium withdrawal in bipolar patients: associations with mania. *Br J Psychiatry* **170**, 426-430 (1997).

100. Yurgelun-Todd, D. et al. fMRI during affect discrimination in bipolar affective disorder. *Bipolar Disord* **2**, 237-248 (2000).

101. Pavuluri, M.N., O'Connor, M.M., Harral, E. & Sweeney, J.A. Affective neural circuitry during facial emotion processing in pediatric bipolar disorder. *Biol Psychiatry* **62**, 158-167 (2007).

102. Pavuluri, M.N., O'Connor, M.M., Harral, E.M. & Sweeney, J.A. An fMRI study of the interface between affective and cognitive neural circuitry in pediatric

bipolar disorder. *Psychiatry Res* **162**, 244-255 (2008).

103. Foland, L.C. et al. Evidence for deficient modulation of amygdala response by prefrontal cortex in bipolar mania. *Psychiatry Res* **162**, 27-37 (2008).

104. Rich, B.A. et al. Limbic hyperactivation during processing of neutral facial expressions in children with bipolar disorder. *Proc Natl Acad Sci U S A* **103**, 8900-8905 (2006).

105. Lawrence, N.S. et al. Subcortical and ventral prefrontal cortical neural responses to facial expressions distinguish patients with bipolar disorder and major depression. *Biol Psychiatry* **55**, 578-587 (2004).

106. Lennox, B., Jacob, R., Calder, A., Lupson, V. & Bullmore, E. Behavioural and neurocognitive responses to sad facial affect are attenuated in patients with mania. *Psychol Med* **34**, 795-802 (2004).

107. Altshuler, L. et al. Increased amygdala activation during mania: a functional magnetic resonance imaging study. *Am J Psychiatry* **162**, 1211-1213 (2005).

108. Dickstein, D.P. et al. Neural activation during encoding of emotional faces in pediatric bipolar disorder. *Bipolar Disord* **9**, 679-692 (2007).

109. Malhi, G.S. et al. Is a lack of disgust something to fear? A functional magnetic resonance imaging facial emotion recognition study in euthymic bipolar disorder patients. *Bipolar Disord* **9**, 345-357 (2007).

110. Chen, C. et al. Explicit and implicit facial affect recognition in manic and depressed States of bipolar disorder: a functional magnetic resonance imaging study. *Biol Psychiatry* **59**, 31-39 (2006).

111. Rich, B.A. et al. Neural connectivity in children with bipolar disorder: impairment in the face emotion processing circuit. *J Child Psychol Psychiatry* **49**, 88-96 (2008).

112. Lagopoulos, J. & Malhi, G.S. A functional magnetic resonance imaging study of emotional Stroop in euthymic bipolar disorder. *Neuroreport* **18**, 1583-1587 (2007).

113. Malhi, G.S., Lagopoulos, J., Sachdev, P.S., Ivanovski, B. & Shnier, R. An emotional Stroop functional MRI study of euthymic bipolar disorder. *Bipolar Disord* **7 Suppl 5**, 58-69 (2005).

114. Adler, C.M., Holland, S.K., Schmithorst, V., Tuchfarber, M.J. & Strakowski, S.M. Changes in neuronal activation in patients with bipolar disorder during performance of a working memory task. *Bipolar Disord* **6**, 540-549 (2004).

115. Monks, P.J. et al. A functional MRI study of working memory task in euthymic bipolar disorder: evidence for task-specific dysfunction. *Bipolar Disord* **6**, 550-564 (2004).

116. Frangou, S., Kington, J., Raymont, V. & Shergill, S.S. Examining ventral and dorsal prefrontal function in bipolar disorder: A functional magnetic resonance imaging study. *Eur Psychiatry* (2007).

117. Frangou, S. The Maudsley Bipolar Disorder Project. *Epilepsia* **46**, 19-25 (2005).

118. Haldane, M. et al. Changes in brain activation during working memory and facial recognition tasks in patients with bipolar disorder with Lamotrigine monotherapy. *Eur Neuropsychopharmacol* **18**, 48-54 (2008).

119. Drapier, D. et al. Genetic Liability for Bipolar Disorder Is Characterized by Excess Frontal Activation in Response to a Working Memory Task. *Biol Psychiatry* **64**, 513-520 (2008).

120. Myers, J. & Well, A. *Research design and statistical analysis*. (L. Erlbaum Associates: Hillsdale N.J., 1995).

121. Lagopoulos, J., Ivanovski, B. & Malhi, G.S. An event-related functional MRI study of working memory in euthymic bipolar disorder. *J Psychiatry Neurosci* **32**, 174-184 (2007).

122. Bechara, A., Damasio, H., Tranel, D. & Anderson, S. Dissociation Of working memory from decision making within the human prefrontal cortex. *J Neurosci* **18**, 428-437 (1998).

123. Koechlin, E., Basso, G., Pietrini, P., Panzer, S. & Grafman, J. The role of the anterior prefrontal cortex in human cognition. *Nature* **399**, 148-151 (1999).

124. Koechlin, E. & Hyafil, A. Anterior prefrontal function and the limits of human decision-making. *Science* **318**, 594-598 (2007).

125. Malhi, G.S. et al. Reduced activation to implicit affect induction in euthymic bipolar patients: an fMRI study. *J Affect Disord* **97**, 109-122 (2007).

126. Thirion, B. et al. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage* **35**, 105-120 (2007).

127. Basser, P., Pajevic, S., Pierpaoli, C., Duda, J. & Aldroubi, A. In vivo fiber tractography using DT-MRI data. *Magn Reson Med* **44**, 625-632 (2000).

128. Jones, D.K. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn Reson Med* **49**, 7-12 (2003).

129. Adler, C.M. et al. Abnormal frontal white matter tracts in bipolar disorder: a diffusion tensor imaging study. *Bipolar Disord* **6**, 197-203 (2004).

130. Haznedar, M. et al. Fronto-thalamo-striatal gray and white matter volumes and anisotropy of their connections in bipolar spectrum illnesses. *Biol Psychiatry* **57**, 733-742 (2005).

131. Beyer, J.L. et al. Cortical white matter microstructural abnormalities in bipolar disorder. *Neuropsychopharmacology* **30**, 2225-2229 (2005).

132. Regenold, W.T. et al. Diffusion-weighted magnetic resonance imaging of white matter in bipolar disorder: a pilot study. *Bipolar Disorders* **8**, 188-195 (2006).

133. Yurgelun-Todd, D.A., Silveri, M.M., Gruber, S.A., Rohan, M.L. & Pimentel, P.J. White matter abnormalities observed in bipolar disorder: a diffusion tensor imaging study. *Bipolar Disord* **9**, 504-512 (2007).

134. Adler, C.M. et al. Evidence of White Matter Pathology in Bipolar Disorder Adolescents Experiencing Their First Episode of Mania: A Diffusion Tensor Imaging Study. *Am J Psychiatry* **163**, 322-324 (2006).

135. Frazier, J.A. et al. White matter abnormalities in children with and at risk for bipolar disorder. *Bipolar Disorders* **9**, 799-809 (2007).

136. Parry, P. & Allison, S. Pre-pubertal paediatric bipolar disorder: a controversy from America. *Australasian Psychiatry* **16**, 80 - 84 (2008).

137. Houenou, J. et al. Increased white matter connectivity in euthymic bipolar patients: diffusion tensor tractography between the subgenual cingulate and the amygdalo-hippocampal complex. *Mol Psychiatry* **12**, 1001-1010 (2007).

138. Bruno, S., Cercignani, M. & Ron, M.A. White matter abnormalities in bipolar disorder: a voxel-based diffusion tensor imaging study. *Bipolar Disorders* **10**, 460-468 (2008).

139. Wang, F. et al. Abnormal Corpus Callosum Integrity in Bipolar Disorder: A Diffusion Tensor Imaging Study. *Biol Psychiatry* **64**, 730-733 (2008).

140. Wang, F. et al. Abnormal anterior cingulum integrity in bipolar disorder determined through diffusion tensor imaging. *Br J Psychiatry* **193**, 126-129 (2008).

141. Smith, S.M. et al. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nat Protoc* **2**, 499-503 (2007).

142. Smith, S.M. et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487-1505 (2006).

143. McIntosh, A.M. et al. White Matter Tractography in Bipolar Disorder and Schizophrenia. *Biol Psychiatry* (2008).

144. Versace, A. et al. Elevated Left and Reduced Right Orbitomedial Prefrontal Fractional Anisotropy in Adults With Bipolar Disorder Revealed by Tract-Based Spatial Statistics. *Arch Gen Psychiatry* **65**, 1041-1052 (2008).

145. Shaltiel, G., Dalton, E.C., Belmaker, R.H., Harwood, A.J. & Agam, G. Specificity of mood stabilizer action on neuronal growth cones. *Bipolar Disord* **9**, 281-289 (2007).

146. Chen, G. & Manji, H.K. The extracellular signal-regulated kinase pathway: an emerging promising target for mood stabilizers. *Curr Opin Psychiatry* **19**, 313-323 (2006).

147. Kim, Y.R. et al. Functional MRI of delayed chronic lithium treatment in rat focal cerebral ischemia. *Stroke* **39**, 439-447 (2008).

148. Joffe, R.T., MacDonald, C. & Kutcher, S.P. Lack of differential cognitive effects of lithium and carbamazepine in bipolar affective disorder. *J Clin Psychopharmacol* **8**, 425-8 (1988).

149. Gualtieri, C. & Johnson, L.G. Comparative neurocognitive effects of 5 psychotropic anticonvulsants and lithium. *MedGenMed* **8**, 46 (2006).

150. Muller-Oerlinghausen, B., Hamann, S., Herrmann, W. & Kropf, D. Effects of lithium on vigilance, psychomotoric performance and mood. *Pharmakopsychiatr Neuropsychopharmakol* **12**, 388-396 (1979).

151. Muller-Oerlinghausen, B., Bauer, H., Girke, W., Kanowski, S. & Goncalves, N. Impairment of vigilance and performance under lithium-treatment. Studies in patients and normal volunteers. *Pharmakopsychiatr Neuropsychopharmakol* **10**, 67-78 (1977).

152. Glue, P., Nutt, D., Cowen, P. & Broadbent, D. Selective effect of lithium on cognitive performance in man. *Psychopharmacology (Berl)* **91**, 109-111 (1987).

153. Stip, E., Dufresne, J., Lussier, I. & Yatham, L. A double-blind, placebo-controlled study of the effects of lithium on cognition in healthy subjects: mild and selective effects on learning. *J Affect Disord* **60**, 147-157 (2000).

154. Weingartner, H., Rudorfer, M. & Linnoila, M. Cognitive effects of lithium treatment in normal volunteers. *Psychopharmacology (Berl)* **86**, 472-474 (1985).

155. Meador, K.J. et al. Comparative cognitive effects of phenobarbital, phenytoin, and valproate in healthy adults. *Neurology* **45**, 1494-1499 (1995).

156. Foland, L.C. et al. Increased volume of the amygdala and hippocampus in bipolar patients treated with lithium. *Neuroreport* **19**, 221-224 (2008).

157. Yucel, K. et al. Bilateral hippocampal volume increase in patients with bipolar disorder and short-term lithium treatment. *Neuropsychopharmacology* **33**, 361-367 (2008).

158. Bearden, C.E. et al. Greater cortical gray matter density in lithium-treated patients with bipolar disorder. *Biol Psychiatry* **62**, 7-16 (2007).

159. Bell, E.C., Willson, M.C., Wilman, A.H., Dave, S. & Silverstone, P.H. Differential effects of chronic lithium and valproate on brain activation in healthy volunteers. *Hum Psychopharmacol* **20**, 415-424 (2005).

160. Caligiuri, M.P. et al. An fMRI study of affective state and medication on cortical and subcortical brain regions during motor performance in bipolar disorder. *Psychiatry Res* **123**, 171-182 (2003).

161. Jogia, J., Haldane, M., Cobb, A., Kumari, V. & Frangou, S. Pilot investigation of the changes in cortical activation during facial affect recognition with lamotrigine monotherapy in bipolar disorder. *Br J Psychiatry* **192**, 197-201 (2008).

162. Gaillard, W., Zeffiro, T., Fazilat, S., DeCarli, C. & Theodore, W. Effect of valproate on cerebral metabolism and blood flow: an 18F-2-deoxyglucose and 15O water positron emission tomography study. *Epilepsia* **37**, 515-521 (1996).

163. Leiderman, D., Balish, M., Bromfield, E. & Theodore, W. Effect of valproate on human cerebral glucose metabolism. *Epilepsia* **32**, 417-422 (1991).

164. Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch Gen Psychiatry* **4**, 561-71 (1961).

165. Altman, E.G., Hedeker, D., Peterson, J.L. & Davis, J.M. The Altman Self-Rating Mania Scale. *Biol Psychiatry* **42**, 948-55 (1997).

166. Hamilton, M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* **23**, 56-62 (1960).

167. Young, R., Biggs, J., Ziegler, V. & Meyer, D. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* **133**, 429-435 (1978).

168. Great Britain Office of Population, C. & Surveys *Standard occupational classification*. (H.M.S.O.: London, 1991).

169. Annett, M. A classification of hand preference by association analysis. *Br J Psychol* **61**, 303-321 (1970).

170. Hatta, T. Handedness and the brain: a review of brain-imaging techniques. *Magn Reson Med Sci* **6**, 99-9112 (2007).

171. Huber, H. [On the problem of variance-analytical reliability estimation in correlated measurement errors]. *Arch Gesamte Psychol* **119**, 288-296 (1967).

172. White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrics* **48**, 817-838 (1980).

173. Williams, R. A note on robust variance estimation for cluster-correlated data. *Biometrics* **56**, 645-646 (2000).

174. Schulze, K. et al. The relationship between eye movement and brain structural abnormalities in patients with schizophrenia and their unaffected relatives. *Journal of Psychiatric Research* **40**, 589-598 (2006).

175. Norwitz, E.R., Edusa, V. & Park, J.S. Maternal Physiology and Complications of Multiple Pregnancy. *Seminars in Perinatology* **29**, 338-348 (2005).

176. Voracek, M. & Haubner, T. Twin-singleton differences in intelligence: a meta-analysis. *Psychol Rep* **102**, 951-962 (2008).

177. Kieseppa, T. et al. Memory and verbal learning functions in twins with bipolar-I disorder, and the role of information-processing speed. *Psychol Med* **35**, 205-215 (2005).

178. Arts, B., Jabben, N., Krabbendam, L. & van Os, J. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychol Med* 1-15 (2007).

179. Weishaupt, D., Koechli, V.D. & Marincek, B. *How does MRI work? An Introduction to the Physics and Function of Magnetic Resonance Imaging*. (Springer: Berlin, 2006).

180. Sternberg, S. High-speed scanning in human memory. *Science* **153**, 652-4 (1966).

181. Baddeley, A. & Hitch Working memory. *The psychology of learning and motivation: Advances in research and theory* **8**, 47-89 (1974).

182. Gevins, A. & Cutillo, B. Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography and Clinical Neurophysiology* **87**, 128-143 (1993).

183. Owen, A.M., McMillan, K.M., Laird, A.R. & Bullmore, E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum Brain Mapp* **25**, 46-59 (2005).

184. Honey, G.D., Bullmore, E.T. & Sharma, T. Prolonged Reaction Time to a Verbal Working Memory Task Predicts Increased Power of Posterior Parietal Cortical Activation*1. *NeuroImage* **12**, 495-503 (2000).

185. Honey, G.D. et al. Effects of Verbal Working Memory Load on Corticocortical Connectivity Modeled by Path Analysis of Functional Magnetic Resonance Imaging Data. *NeuroImage* **17**, 573-582 (2002).

186. Amaro, J. & Barker, G.J. Study design in fMRI: Basic principles. *Brain and Cognition* **60**, 220-232 (2006).

187. Mechelli, A., Henson, R.N.A., Price, C.J. & Friston, K.J. Comparing event-related and epoch analysis in blocked design fMRI. *NeuroImage* **18**, 806-810 (2003).

188. Harkavy-Friedman, J.M. et al. Are BPI and BPII suicide attempters distinct neuropsychologically? *Journal of Affective Disorders* **94**, 255-259 (2006).

189. Sullivan, E.V. & Pfefferbaum, A. Neurocircuitry in alcoholism: a substrate of disruption and repair. *Psychopharmacology (Berl)* **180**, 583-594 (2005).

190. Owen, A.M. et al. Functional organization of spatial and nonspatial working memory processing within the human lateral frontal cortex. *Proc Natl Acad Sci U S A* **95**, 7721-6 (1998).

191. Mumford, J.A. & Nichols, T.E. Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *NeuroImage* **39**, 261-268 (2008).

192. Voss, H.U., Zevin, J.D. & McCandliss, B.D. Functional MR Imaging at 3.0 T versus 1.5 T: A Practical Review. *Neuroimaging Clinics of North America* **16**, 285-297 (2006).

193. Bora, E. et al. The effect of previous psychotic mood episodes on cognitive impairment in euthymic bipolar patients. *Bipolar Disord* **9**, 468-477 (2007).

194. Martinez-Aran, A. et al. Neurocognitive impairment in bipolar patients with and without history of psychosis. *J Clin Psychiatry* **69**, 233-239 (2008).

195. Selva, G. et al. Bipolar I patients with and without a history of psychotic symptoms: do they differ in their cognitive functioning? *J Psychiatr Res* **41**, 265-272 (2007).

196. Friston, K. et al. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6**, 218-229 (1997).

197. Logothetis, N.K. & Pfeuffer, J. On the nature of the BOLD fMRI contrast

mechanism. *Magn Reson Imaging* **22**, 1517-1531 (2004).

198. Logothetis, N.K. & Wandell, B.A. Interpreting the BOLD Signal. *Annual Review of Physiology* **66**, 735-769 (2004).

199. Logothetis, N.K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869-878 (2008).

200. McIntosh, A.M. et al. The effects of a neuregulin 1 variant on white matter density and integrity. *Mol Psychiatry* **13**, 1054-1059 (2007).

201. Georgieva, L. et al. Support for Neuregulin 1 as a Susceptibility Gene for Bipolar Disorder and Schizophrenia. *Biological Psychiatry* **64**, 419-427 (2008).

202. Prata, D.P. et al. An association study of the neuregulin 1 gene, bipolar affective disorder and psychosis. *Psychiatric Genetics* **19**, 113-116 (2009).

203. Goes, F.S. et al. Family-based association study of Neuregulin 1 with psychotic bipolar disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet* **150B**, 693-702 (2009).

204. Palo, O.M. et al. Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments. *Hum. Mol. Genet.* **16**, 2517-2528 (2007).

205. Thomson, P.A. et al. Association between the TRAX//DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Mol Psychiatry* **10**, 657-668 (2005).

206. Schosser, A. et al. Association of DISC1 and TSNAX genes and affective disorders in the depression case-control (DeCC) and bipolar affective case-control (BACCS) studies. *Mol Psychiatry* (2009).at <http://dx.doi.org/10.1038/mp.2009.21>

207. Hashimoto, R. et al. Impact of the DISC1 Ser704Cys polymorphism on risk for major depression, brain morphology and ERK signaling. *Hum. Mol. Genet.* **15**, 3024-3033 (2006).

208. Mao, Y. et al. Disrupted in Schizophrenia 1 Regulates Neuronal Progenitor Proliferation via Modulation of GSK3[beta]/[beta]-Catenin Signaling. *Cell* **136**, 1017-1031 (2009).

209. Aubry, J., Schwald, M., Ballmann, E. & Karege, F. Early effects of mood stabilizers on the Akt/GSK-3beta signaling pathway and on cell survival and proliferation. *Psychopharmacology (Berl.)* **205**, 419-429 (2009).

210. Machado-Vieira, R., Manji, H.K. & Zarate, C.A. The role of lithium in the

treatment of bipolar disorder: convergent evidence for neurotrophic effects as a unifying hypothesis. *Bipolar Disord* **11 Suppl 2**, 92-109 (2009).

211. Morris, J.A., Kandpal, G., Ma, L. & Austin, C.P. DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. *Hum. Mol. Genet.* **12**, 1591-1608 (2003).

212. Barnett, J. & Smoller, J. The genetics of bipolar disorder. *Neuroscience* **In Press, Corrected Proof**,

213. Derek Kenton Jones, S.C.R.W. Isotropic resolution diffusion tensor imaging with whole brain acquisition in a clinically acceptable time. *Human Brain Mapping* **15**, 216-230 (2002).

214. Basser, P.J. & Pierpaoli, C. Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI. *Journal of Magnetic Resonance, Series B* **111**, 209-219 (1996).

215. Good, C. et al. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* **14**, 21-36 (2001).

216. Jones, D.K., Symms, M.R., Cercignani, M. & Howard, R.J. The effect of filter size on VBM analyses of DT-MRI data. *Neuroimage* **26**, 546-554 (2005).

217. Bullmore, E. et al. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans Med Imaging* **18**, 32-42 (1999).

218. Carrew, J., Whaba, G. & Basser, P. Implications of heterogeneous of tensor-derived quantities for group comparisons. *Proc. Int.l Soc. Mag. Reson. Med.* 1598 (2007).

219. Mori, S. & van Zijl, P. Human white matter atlas. *Am J Psychiatry* **164**, 1005 (2007).

220. Bruno, S., Cercignani, M. & Ron, M.A. White matter abnormalities in bipolar disorder: a voxel-based diffusion tensor imaging study. *Bipolar Disorders* **10**, 460-468 (2008).

221. Vernooij, M. et al. White matter atrophy and lesion formation explain the loss of structural integrity of white matter in aging. *Neuroimage* (2008).

222. Yoon, B., Shim, Y., Lee, K., Shon, Y. & Yang, D. Region-specific changes of cerebral white matter during normal aging: a diffusion-tensor analysis. *Arch Gerontol Geriatr* **47**, 129-138 (2008).

223. Hasan, K.M. et al. Diffusion tensor quantification of the human midsagittal corpus callosum subdivisions across the lifespan. *Brain Res* **1227**, 52-67 (2008).

224. Makris, N. et al. Segmentation of Subcomponents within the Superior Longitudinal Fascicle in Humans: A Quantitative, In Vivo, DT-MRI Study. *Cereb. Cortex* **15**, 854-869 (2005).

225. Petrides, M. & Pandya, D. Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J Comp Neurol* **228**, 105-116 (1984).

226. Mori, S., Crain, B., Chacko, V. & van Zijl, P. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol* **45**, 265-269 (1999).

227. Makris, N. et al. Morphometry of in vivo human white matter association pathways with diffusion-weighted magnetic resonance imaging. *Ann Neurol* **42**, 951-962 (1997).

228. Catani, M., Howard, R.J., Pajevic, S. & Jones, D.K. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* **17**, 77-94 (2002).

229. Preuss, T. & Goldman-Rakic, P. Connections of the ventral granular frontal cortex of macaques with perisylvian premotor and somatosensory areas: anatomical evidence for somatic representation in primate frontal association cortex. *J Comp Neurol* **282**, 293-316 (1989).

230. Petrides, M. & Pandya, D. Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur J Neurosci* **11**, 1011-1036 (1999).

231. Banich, M. Interhemispheric interaction: Mechanisms of unified processing. *Brain Asymmetry* 427-450 (1995).

232. Bloom, J.S. & Hynd, G.W. The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychol Rev* **15**, 59-71 (2005).

233. Witelson, S. Hand and sex differences in the isthmus and genu of the human corpus callosum. A postmortem morphological study. *Brain* **112 (Pt 3)**, 799-835 (1989).

234. Hofer, S. & Frahm, J. Topography of the human corpus callosum revisited--Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *NeuroImage* **32**, 989-994 (2006).

235. Jea, A. et al. Corpus callosotomy in children and the disconnection syndromes: a

review. *Childs Nerv Syst* **24**, 685-692 (2008).

236. Paul, L.K. et al. Agenesis of the corpus callosum: genetic, developmental and functional aspects of connectivity. *Nat Rev Neurosci* **8**, 287-299 (2007).

237. Wahlsten, D. Deficiency of corpus callosum varies with strain and supplier of the mice. *Brain Research* **239**, 329-347 (1982).

238. Clapcote, S.J. & Roder, J.C. Deletion Polymorphism of Disc1 Is Common to All 129 Mouse Substrains: Implications for Gene-Targeting Studies of Brain Function. *Genetics* **173**, 2407-2410 (2006).

239. Koike, H., Arguello, P.A., Kvajo, M., Karayiorgou, M. & Gogos, J.A. Disc1 is mutated in the 129S6/SvEv strain and modulates working memory in mice. *Proceedings of the National Academy of Sciences* **103**, 3693-3697 (2006).

240. Mandonnet, E., Nouet, A., Gatignol, P., Capelle, L. & Duffau, H. Does the left inferior longitudinal fasciculus play a role in language? A brain stimulation study. *Brain* **130**, 623-629 (2007).

241. Catani, M., Jones, D.K., Donato, R. & Ffytche, D.H. Occipito-temporal connections in the human brain. *Brain* **126**, 2093-2107 (2003).

242. Phillips, M.L. The neural basis of mood dysregulation in bipolar disorder. *Cognit Neuropsychiatry* **11**, 233-249 (2006).

243. Lembke, A. & Ketter, T.A. Impaired Recognition of Facial Emotion in Mania. *Am J Psychiatry* **159**, 302-304 (2002).

244. Friston, K., Harrison, L. & Penny, W. Dynamic causal modelling. *Neuroimage* **19**, 1273-1302 (2003).

245. Rijsdijk, F.V. & Sham, P.C. Analytic approaches to twin data using structural equation models. *Brief Bioinform* **3**, 119-33 (2002).

246. Nestor, P.G. et al. Neuropsychological correlates of diffusion tensor imaging in schizophrenia. *Neuropsychology* **18**, 629-637 (2004).

247. Nestor, P.G. et al. Neuropsychological disturbance in schizophrenia: a diffusion tensor imaging study. *Neuropsychology* **22**, 246-254 (2008).

248. Szeszko, P.R. et al. Clinical and neuropsychological correlates of white matter abnormalities in recent onset schizophrenia. *Neuropsychopharmacology* **33**, 976-984 (2008).

249. Lohoff, F. et al. Confirmation of association between the Val66Met polymorphism in the brain-derived neurotrophic factor (BDNF) gene and bipolar

I disorder. *Am J Med Genet B Neuropsychiatr Genet* **139B**, 51-53 (2005).

250. Buonanno, A. et al. Neuregulins and neuronal plasticity: possible relevance in schizophrenia. *Novartis Found Symp* **289**, 165-177 (2008).

251. Gassmann, M. & Lemke, G. Neuregulins and neuregulin receptors in neural development. *Curr Opin Neurobiol* **7**, 87-92 (1997).

252. Medland, S.E. & Keller, M.C. Modeling extended twin family data II: power associated with different family structures. *Twin Res Hum Genet* **12**, 19-25 (2009).

253. Ladd, M.E. High-field-strength magnetic resonance: potential and limits. *Top Magn Reson Imaging* **18**, 139-152 (2007).

254. Descoteaux, M., Deriche, R., Knösche, T.R. & Anwander, A. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Trans Med Imaging* **28**, 269-286 (2009).

255. Craddock, N., Jones, I., Kirov, G. & Jones, L. The Bipolar Affective Disorder Dimension Scale (BADDS)--a dimensional scale for rating lifetime psychopathology in bipolar spectrum disorders. *BMC Psychiatry* **4**, 19 (2004).

8. Appendices:

Appendix A: consists only of novel scripts. Some of the scripts call on standard functions from the SPM library that have been specifically altered for the purposes in hand. These scripts, where called are generally in the form spm_scriptname_FK. If required, these modifed scripts are available on request from the author.

Appendix B: provides a list of the acronyms used in the current study.

8.1. Appendix A: MATLAB Scripts

Appendix A1: N-back master script

```
%%NBACK MASTER SCRIPT

% Developed by Fergus Kane, 2007.

% Calls the following functions:

    %NBACK_readsubjectIDs – same as ReadSubjectID.m
    %NBACK_RFAnalyse – Analysed all subjects response files, generates vectors
    %NBACK_specify – specify model for each subject using results of RFAnalyse
    %NBACK_estimate
    %NBACK_AddContrasts

%Set Paths (may not all be necessary - can't remember!)

        path ('/biofs/software/system/spm/spm5/toolbox/marsbar/spm5', path)
        path ('/biofs/software/system/spm/spm5/', path)
    path ('/home/spnefek/matlab/NBACK/', path)


%CALL SPM

    clear all

        spm fmri
        spm_defaults
        global defaults

%SET UP VARIABLES


    %set base directory containing subjects directories
        v.basedir = '/home/spnefek/NBACK/Standard3Condition/';
    %set directory of template job file
        v.jobsfiledir = '/home/spnefek/matlab/NBACK/';
    %set template job file for model specification
        v.templatefile = 'NBACK_model_6cond.mat';
    %set directory holding subjects text file
        v.subjectsdir = '/home/spnefek/matlab/NBACK/responsefiles/';
    %set directory holding responsefiles

        v.rfd='/home/spnefek/matlab/NBACK/responsefiles/';
    %set subjects text file name
        v.fn = 'AllSubjects.prn';
    %set constrasts job file
        v.contrastfile='/home/spnefek/NBACK/Smooth/contrasttemp.mat'


%All functions (not really functions becuae of the way they are set up, more linked
%scripts) have multiple subject loops built in, these could be
%extracted and placed in this masterscript.

%Response file directory must contain:

    %a. Response files in format TBNB[SubjectID].TXT
    %b. Subjects file
    %c. Condition vectors: 1back.prn, 2back.prn, 3back.prn, isitx.prn


%0. LOAD THE SUBJECTS FILE, This is required for the loops:

    %   Reads subjects file with two columns, first is the subject directory
    %   name and second is the subject ID (white space seperated).



    v.SubjectIDs = NBACK_readsubjectID(v.subjectsdir, v.fn)
```

```
%1. ANALYSE RESPONSE FILES

    %Run Analysis and feed back necessary variables

    [BehavData, RespVect, ExportData] = NBACK_RFAnalyse(v.rfd, v.fn);

    v.str='Analysis Complete';


 %2. RUN Model Specification - Requires output variables from Analyse Step

     NBACK_specify(BehavData, RespVect, v.basedir, v.jobsfiledir, v.templatefile,
v.SubjectIDs);
     v.str='Specification Complete';


% 3. RUN Model Estimation

     NBACK_estimate(v.basedir, v.SubjectIDs);
     v.str='Estimation Complete';



% %4. ADD Contrasts
     clear v.SubjectIDs
     v.fn = 'AllSubjects.prn';

     v.SubjectIDs = NBACK_readsubjectID(v.subjectsdir, v.fn)
     NBACK_AddContrasts(v.contrastfile,v.basedir, v.SubjectIDs);
     v.str='Contrasts Added';
```

Appendix A2: RFAnalyse.m

```
function [BehavData, RespVect, ExportData] = NBACK_RFAnalyse(rfd, fn)
%SCRIPT FOR Analysing NBACK Data Files - Fergus Kane

%Warning, there may be problems if the response files are dodgy... in case
%of error, check the loop to see which response file has crashed.

%The algoriths in the file are a bit messy and could be significantly
%tidied up - but it works!

% Go to Project Directory

    cd(rfd);

%read condition vectors, this sets up the template from the current study,
%which is the same for all subjects
    v.fn = '1back.prn';
    v.fid=fopen(v.fn);
    Data.CVectors.oneback=textscan(v.fid, '%n');
    fclose(v.fid);

    v.fn = '2back.prn';
    v.fid=fopen(v.fn);
    Data.CVectors.twoback=textscan(v.fid, '%n');
    fclose(v.fid);

    v.fn = '3back.prn';
    v.fid=fopen(v.fn);
    Data.CVectors.threeback=textscan(v.fid, '%n');
    fclose(v.fid);

    v.fn = 'isitx.prn';
    v.fid=fopen(v.fn);
    Data.CVectors.isitx=textscan(v.fid, '%n');
    fclose(v.fid);

    clear fid

%read subject IDs, with twinbip first (scanner ID) and standard study ID
%second.

    fid=fopen(fn);
    Data.SubjectIDs=textscan(fid, '%n %n');
    fclose(fid);

%Set up loop to run analysis for all subjects

    for i=1:size(Data.SubjectIDs{1,1},1);

        v.id=Data.SubjectIDs{1,1}(i);
        v.stid=Data.SubjectIDs{1,2}(i);

     %set response file name
        v.rfn = strcat('TBNB', num2str(v.stid), '.TXT');
        %just to check
        BehavData(v.id).rfn=v.rfn;
        BehavData(v.id).stid=v.stid;

            %read response file
            v.fid=fopen(v.rfn);
            if v.fid == -1;
                v.warning = strcat('Can not open ', v.rfn);
                warning(v.warning);
            else
            end

        %skip 1st line and 1st character of second line
            v.seekok = fseek(v.fid, 78, 'bof');
            if v.seekok == -1;
                warning('failed to seek');
            else
            end
            clear v.seekok;


        %reads file into matrix Data.C, number, text, text, number, number, number
```

```
            Data.C=textscan(v.fid,'%n %13c %13c %n %n %n');
            fclose(v.fid);

           %convert Matrix Data.C, column 3  (response column) into cell array
            Data.C{1,3} =cellstr(Data.C{1,3});
        %make copy of response column with all 'nones' replaced with 0s
            Data.Responses = regexprep(Data.C{1,3}, 'None', '0');
        %make copy of response time column
            Data.RespT = Data.C{1,4};
        %make copy of Data.Stimuli column
            Data.Stimuli = Data.C{1,6};


        % create binary list of Data.Stimuli, 'Data.Stimuli2', where 1 = expected
response
        % and 0 represents no expected response

        %(first get array length)
            v.Count=length(Data.Stimuli);

            for j = 1:v.Count;
                if Data.Stimuli(j,1)==1;
                    Data.Stimuli2(j,1)=1;
                else Data.Stimuli2(j,1)=0;
            end
            end



        %RESPONSE MATRIX Data.CREATION

        %create matrix correctresp: column1 = Counter,
        %column1=correct response times
        %column2=

            y=1;
            for j = 1:v.Count;
                    if (Data.Stimuli2(j,1)==1 && Data.RespT(j,1)>0);
                    RespVect(v.id).CorrectResp(y,1) = Data.C{1,1}(j,1);
                    RespVect(v.id).CorrectResp(y,2) = Data.C{1,4}(j,1);
                    RespVect(v.id).CorrectResp(y,3) = Data.C{1,5}(j,1);
                    y=y+1;
                else
                 end
            end

            if isempty(RespVect(v.id).CorrectResp) == 0

BehavData(v.id).CorrectResponses.MeanRT=mean(RespVect(v.id).CorrectResp(1:end,2));

BehavData(v.id).CorrectResponses.NumberResp=size(RespVect(v.id).CorrectResp,1);
            else
            end

            clear y;
            clear j;



        %create matrix falsepositive column1 = Counter
        %column2=correct response times
        %column3=

            y=1;
            for j = 1:v.Count;
                 if (Data.Stimuli2(j,1)<1 && Data.RespT(j,1)>0);
                    RespVect(v.id).FalsePos(y,1) = Data.C{1,1}(j,1);
                    RespVect(v.id).FalsePos(y,2) = Data.C{1,4}(j,1);
                    RespVect(v.id).FalsePos(y,3) = Data.C{1,5}(j,1);
                    y=y+1;
                else
                 end
            end

            if isempty(RespVect(v.id).FalsePos) == 0

BehavData(v.id).FalsePositives.MeanRT=mean(RespVect(v.id).FalsePos(1:end,2));
```

```
BehavData(v.id).FalsePositives.NumberResp=size(RespVect(v.id).FalsePos,1);
            else
            end


            clear y;
            clear j;

        %create matrix falsenegative column1 = Counter
        %column2=correct response times
        %column3=

            y=1;
            for j = 1:v.Count;
                 if (Data.Stimuli2(j,1)==1 && Data.RespT(j,1)==0);
                    RespVect(v.id).FalseNeg(y,1) = Data.C{1,1}(j,1);
                    RespVect(v.id).FalseNeg(y,2) = Data.C{1,4}(j,1);
                    RespVect(v.id).FalseNeg(y,3) = Data.C{1,5}(j,1);
                    y=y+1;
                else
                 end
            end

            if isempty(RespVect(v.id).FalseNeg) == 0

BehavData(v.id).FalseNeg.MeanRT=mean(RespVect(v.id).FalseNeg(1:end,2));
                BehavData(v.id).FalseNeg.NumberResp=size(RespVect(v.id).FalseNeg,1);
            else
            end

            clear y;
            clear j;

        %create matrix allerrors column1 = Counter
        %column2=correct response times
        %column3=

            y=1;
            for j = 1:v.Count;
                 if ((Data.Stimuli2(j,1)==1 && Data.RespT(j,1)==0) |
(Data.Stimuli2(j,1)<1 && Data.RespT(j,1)>0)) ;
                    RespVect(v.id).AllErrors(y,1) = Data.C{1,1}(j,1);
                    RespVect(v.id).AllErrors(y,2) = Data.C{1,4}(j,1);
                    RespVect(v.id).AllErrors(y,3) = Data.C{1,5}(j,1);
                    y=y+1;
                else
                 end
            end

            if isempty(RespVect(v.id).FalseNeg) == 0

BehavData(v.id).AllErrors.MeanRT=mean(RespVect(v.id).AllErrors(1:end,2));
                BehavData(v.id).AllErrors.NumberResp=size(RespVect(v.id).AllErrors,1);
            else
            end

            clear y;
            clear j;



        %create matrix allbuttonpress column1 = Counter
        %column2=correct response times
        %column3=

            y=1;
            for j = 1:v.Count;
                 if ((Data.Stimuli2(j,1)>0 && Data.RespT(j,1)>0) |
(Data.Stimuli2(j,1)<1 && Data.RespT(j,1)>0));
                    RespVect(v.id).Response(y,1) = Data.C{1,1}(j,1);
                    RespVect(v.id).Response(y,2) = Data.C{1,4}(j,1);
                    RespVect(v.id).Response(y,3) = Data.C{1,5}(j,1);
                    y=y+1;
                else
                 end
            end

            if isempty(RespVect(v.id).Response) == 0
```

```
BehavData(v.id).Response.MeanRT=mean(RespVect(v.id).Response(1:end,2));
                BehavData(v.id).Response.NumberResp=size(RespVect(v.id).Response,1);
            else
            end
            clear y;
            clear j;

        %create matrix all responses and false negatives (nointerest)  column1 =
Counter
        %column2=correct response times
        %column3=

            y=1;
            for j = 1:v.Count;
                 if ((Data.Stimuli2(j,1)>0 && Data.RespT(j,1)>0) |
(Data.Stimuli2(j,1)<1 && Data.RespT(j,1)>0) | (Data.Stimuli2(j,1)==1 &&
Data.RespT(j,1)==0));
                    RespVect(v.id).NoInterest(y,1) = Data.C{1,1}(j,1);
                    RespVect(v.id).NoInterest(y,2) = Data.C{1,4}(j,1);
                    RespVect(v.id).NoInterest(y,3) = Data.C{1,5}(j,1);
                    y=y+1;
                else
                 end
            end

            if isempty(RespVect(v.id).NoInterest) == 0

BehavData(v.id).NoInterest.MeanRT=mean(RespVect(v.id).Response(1:end,2));
                BehavData(v.id).NoInterest.NumberResp=size(RespVect(v.id).Response,1);
            else
            end
            clear y;
            clear j;



        %The following matrix creations use the conditionvectors imported earlier.
        %-------------------------------------------------------------------------
        %Data.Create Matrix of Non-Data.Responses for 'is it x'
            x=1;
            for j= 1:length(Data.CVectors.isitx{1});
                y=Data.CVectors.isitx{1,1}(j);
                if (Data.Stimuli2(y)<1 && Data.RespT(y,1)<0.0001);
                    RespVect(v.id).NRisitx(x,1) = Data.C{1,1}(y);
                    RespVect(v.id).NRisitx(x,2) = Data.C{1,4}(y);
                    RespVect(v.id).NRisitx(x,3) = Data.C{1,5}(y);
                    x=x+1;
                else
                end
            end

        %Data.Create Matrix of Data.Correct Data.Responses for 'is it x'
            x=1;
            for j= 1:length(Data.CVectors.isitx{1});
                y=Data.CVectors.isitx{1,1}(j);
                if (Data.Stimuli2(y)==1 && Data.RespT(y,1)>0);
                    RespVect(v.id).CRisitx(x,1) = Data.C{1,1}(y);
                    RespVect(v.id).CRisitx(x,2) = Data.C{1,4}(y);
                    RespVect(v.id).CRisitx(x,3) = Data.C{1,5}(y);
                    x=x+1;
                else
                end
            end

            if isempty(RespVect(v.id).CRisitx) == 0
                BehavData(v.id).CRisitx.MeanRT=mean(RespVect(v.id).CRisitx(1:end,2));
                BehavData(v.id).CRisitx.NumberResp=size(RespVect(v.id).CRisitx,1);
            else
            end

        %Data.Create Matrix of Non-Data.Responses for 'oneback'
            x=1;
            for j= 1:length(Data.CVectors.oneback{1});
                y=Data.CVectors.oneback{1,1}(j);
                if (Data.Stimuli2(y)<1 && Data.RespT(y,1)<0.0001);
                    RespVect(v.id).NRoneback(x,1) = Data.C{1,1}(y);
                    RespVect(v.id).NRoneback(x,2) = Data.C{1,4}(y);
                    RespVect(v.id).NRoneback(x,3) = Data.C{1,5}(y);
```

```
                x=x+1;
            else
            end
        end



    %Data.Create Matrix of Data.Correct Data.Responses for 'oneback'
        x=1;
        for j= 1:length(Data.CVectors.oneback{1});
            y=Data.CVectors.oneback{1,1}(j);
            if (Data.Stimuli2(y)==1 && Data.RespT(y,1)>0)
                RespVect(v.id).CRoneback(x,1) = Data.C{1,1}(y);
                RespVect(v.id).CRoneback(x,2) = Data.C{1,4}(y);
                RespVect(v.id).CRoneback(x,3) = Data.C{1,5}(y);
                x=x+1;
            else
            end
        end

        if isempty(RespVect(v.id).CRoneback) == 0

BehavData(v.id).CRoneback.MeanRT=mean(RespVect(v.id).CRoneback(1:end,2));
            BehavData(v.id).CRoneback.NumberResp=size(RespVect(v.id).CRoneback,1);
        else
        end

    %Data.Create Matrix of Non-Data.Responses for 'twoback'
        x=1;
        for j= 1:length(Data.CVectors.twoback{1});
            y=Data.CVectors.twoback{1,1}(j);
            if (Data.Stimuli2(y)<1 && Data.RespT(y,1)<0.0001);
                RespVect(v.id).NRtwoback(x,1) = Data.C{1,1}(y);
                RespVect(v.id).NRtwoback(x,2) = Data.C{1,4}(y);
                RespVect(v.id).NRtwoback(x,3) = Data.C{1,5}(y);
                x=x+1;
            else
            end
        end

    %Data.Create Matrix of Data.Correct Data.Responses for 'twoback'
        x=1;
        for j= 1:length(Data.CVectors.twoback{1});
            y=Data.CVectors.twoback{1,1}(j);
            if (Data.Stimuli2(y)==1 && Data.RespT(y,1)>0);
                RespVect(v.id).CRtwoback(x,1) = Data.C{1,1}(y);
                RespVect(v.id).CRtwoback(x,2) = Data.C{1,4}(y);
                RespVect(v.id).CRtwoback(x,3) = Data.C{1,5}(y);
                x=x+1;
            else
            end
        end

        if isempty(RespVect(v.id).CRtwoback) == 0

BehavData(v.id).CRtwoback.MeanRT=mean(RespVect(v.id).CRtwoback(1:end,2));
            BehavData(v.id).CRtwoback.NumberResp=size(RespVect(v.id).CRtwoback,1);
        else
        end

    %Data.Create Matrix of Non-Data.Responses for 'threeback'
        x=1;
        for j= 1:length(Data.CVectors.threeback{1});
            y=Data.CVectors.threeback{1,1}(j);
            if (Data.Stimuli2(y)<1 && Data.RespT(y,1)<0.0001);
                RespVect(v.id).NRthreeback(x,1) = Data.C{1,1}(y);
                RespVect(v.id).NRthreeback(x,2) = Data.C{1,4}(y);
                RespVect(v.id).NRthreeback(x,3) = Data.C{1,5}(y);
                x=x+1;
            else
            end
        end


    %Data.Create Matrix of Data.Correct Data.Responses for 'threeback'
        x=1;
        for j= 1:length(Data.CVectors.threeback{1});
            y=Data.CVectors.threeback{1,1}(j);
            if (Data.Stimuli2(y)==1 && Data.RespT(y,1)>0);
```

```
                    RespVect(v.id).CRthreeback(x,1) = Data.C{1,1}(y);
                    RespVect(v.id).CRthreeback(x,2) = Data.C{1,4}(y);
                    RespVect(v.id).CRthreeback(x,3) = Data.C{1,5}(y);
                    x=x+1;
                else
                end
            end

            if isempty(RespVect(v.id).CRthreeback) == 0

BehavData(v.id).CRthreeback.MeanRT=mean(RespVect(v.id).CRthreeback(1:end,2));

BehavData(v.id).CRthreeback.NumberResp=size(RespVect(v.id).CRthreeback,1);
            else
            end

            BehavData(v.id).Analysed='done';

            clear x j y v.Count ans;



    %-----------------------------------------------------------------------

end


    %now generate a cell array with Behavioural Data for export and analysis:
    %Generate Headers
        ExportData{1,1}='TWINBIP';
        ExportData{1,2}='SID';
        ExportData{1,3}='CorResp-Tot';
        ExportData{1,4}='FalseP';
        ExportData{1,5}='FalseN';
        ExportData{1,6}='AllErrors';
        ExportData{1,7}='CorR-ItisX';
        ExportData{1,8}='CorR-1B';
        ExportData{1,9}='CorR-2B';
        ExportData{1,10}='CorR-3B';
        ExportData{1,11}='RT-AllCorr';
        ExportData{1,12}='RT-FP';
        ExportData{1,13}='RT-CRX';
        ExportData{1,14}='RT-CR1';
        ExportData{1,15}='RT-CR2';
        ExportData{1,16}='RT-CR3';

    k=1;
    for i= 1:size(BehavData,2);
        if ~isempty(BehavData(1,i).rfn);
            ExportData{k+1,1}=i;
            if ~isempty(BehavData(1,i).stid)
                ExportData{k+1,2}=BehavData(1,i).stid;
            else
            end
            if ~isempty(BehavData(1,i).CorrectResponses);
                ExportData{k+1,3}=BehavData(1,i).CorrectResponses.NumberResp(1);
            else
            end
            if ~isempty(BehavData(1,i).FalsePositives) ;
                ExportData{k+1,4}=BehavData(1,i).FalsePositives.NumberResp(1);
            else
            end
            if ~isempty(BehavData(1,i).FalseNeg);
                ExportData{k+1,5}=BehavData(1,i).FalseNeg.NumberResp(1);
            else
            end
            if ~isempty(BehavData(1,i).FalseNeg);
                ExportData{k+1,6}=BehavData(1,i).AllErrors.NumberResp(1);
            else
            end
            if ~isempty(BehavData(1,i).CRisitx);
                ExportData{k+1,7}=BehavData(1,i).CRisitx.NumberResp(1);
            else
            end
            if ~isempty(BehavData(1,i).CRoneback);
                ExportData{k+1,8}=BehavData(1,i).CRoneback.NumberResp(1);
            else
            end
            if ~isempty(BehavData(1,i).CRtwoback);
```

```
            ExportData{k+1,9}=BehavData(1,i).CRtwoback.NumberResp(1);
        else
        end
        if ~isempty(BehavData(1,i).CRthreeback);
            ExportData{k+1,10}=BehavData(1,i).CRthreeback.NumberResp(1);
        else
        end

        if ~isempty(BehavData(1,i).CorrectResponses);
            ExportData{k+1,11}=BehavData(1,i).CorrectResponses.MeanRT(1);
        else
        end

        if ~isempty(BehavData(1,i).FalsePositives);
            ExportData{k+1,12}=BehavData(1,i).FalsePositives.MeanRT(1);
        else
        end

        if ~isempty(BehavData(1,i).CRisitx);
            ExportData{k+1,13}=BehavData(1,i).CRisitx.MeanRT(1);
        else
        end
        if ~isempty(BehavData(1,i).CRoneback);
            ExportData{k+1,14}=BehavData(1,i).CRoneback.MeanRT(1);
        else
        end
        if ~isempty(BehavData(1,i).CRtwoback);
            ExportData{k+1,15}=BehavData(1,i).CRtwoback.MeanRT(1);
        else
        end
        if ~isempty(BehavData(1,i).CRthreeback);
            ExportData{k+1,16}=BehavData(1,i).CRthreeback.MeanRT(1);
        else
        end




        k=k+1;
    else

    end
end
clear k i
```

Appendix A3: NBACK_specify.txt

```
function a = NBACK_specify(BehavData, RespVect, basedir, jobsfiledir, templatefile,
SubjectIDs)
%Must be run from masterNBACK, after NBACK_RFAnalyse

%%%SECTION 2.  SPECIFY MODEL
%%%Use template jobs file, modify and run for each subject.

%OPEN and EDIT jobs file
    cd(jobsfiledir);
    load(templatefile);

for i=1:size(SubjectIDs{1,1},1);
%to test
%for i=1:2

    v.jid=SubjectIDs{1,1}(i);
    % v.stid=SubjectIDs{1,2}(i);
        v.imagedir = strcat(basedir, num2str(v.jid), '/');

    BehavData(v.jid).status = 'starting specification';


    %set standard variables
        %jobs{1,1}.stats{1,1}.fmri_spec.timing.units='secs';
        %jobs{1,1}.stats{1,1}.fmri_spec.timing.RT=2;
        %jobs{1,1}.stats{1,1}.fmri_spec.timing.fmri_t=16;
        %jobs{1,1}.stats{1,1}.fmri_spec.timing.fmri_t0=1;

        %jobs{1,1}.stats{1,1}.fmri_spec.bases.hrf.volt = 1;
        %jobs{1,1}.stats{1,1}.fmri_spec.bases.hrf.global = 'none';
        %jobs{1,1}.stats{1,1}.fmri_spec.bases.hrf.mask = '"';
        %jobs{1,1}.stats{1,1}.fmri_spec.bases.hrf.cvi = 'AR(1)';

    %specify temporal derivatives as no derivative [0 0] or derivs [1 0]
        %jobs{1,1}.stats{1,1}.fmri_spec.bases.hrf.derivs = [1 0]

    %set subject specific variables

    %first select image files
        %set filter var
            v.filt = '^swu.*\.img$';

        %output list of files
            v.files=spm_select('List',v.imagedir,v.filt,1);
        %concatanate directory with files
            v.files = strcat(v.imagedir,filesep,v.files);
        %convert char array to cell array
            v.files=cellstr(v.files);
        %set scans
            jobs{1,1}.stats{1,1}.fmri_spec.sess.scans = v.files;

        %set directory
        jobs{1,1}.stats{1,1}.fmri_spec.dir{1,1}=v.imagedir



    %now set up conditions
    %NEED TO DECIDE ON CONDITIONS - This is for 5 conditions edit if more!!
        %names
        jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,1).name='baseline';
        jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,2).name='1back';
        jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,3).name='2back' ;
        jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,4).name='3back'  ;
        jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,5).name='correctresp';
        jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,6).name='errors';
        %durations, temporal modulation and pmod
        %for i=1:5
            %jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,i).duration=0
            %jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,i).tmod=0
            %jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,i).pmod=struct
        %end
        %clear i
```

```
        %assign onsets, where there are no onsets, SPM has a fit, so we
        %need to assign a dummy onset, here we have chosen 0.1

        if ~isempty(RespVect(1,v.jid).NRisitx)

jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,1).onset=RespVect(1,v.jid).NRisitx(1:end,3)
;
        else
            jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,1).onset=[0.1]
        end

        if ~isempty(RespVect(1,v.jid).NRoneback)

jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,2).onset=RespVect(1,v.jid).NRoneback(1:end,
3);
        else
            jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,2).onset=[0.1]
        end

        if ~isempty(RespVect(1,v.jid).NRtwoback)

jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,3).onset=RespVect(1,v.jid).NRtwoback(1:end,
3);
        else
            jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,3).onset=[0.1]
        end

        if ~isempty(RespVect(1,v.jid).NRthreeback)

jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,4).onset=RespVect(1,v.jid).NRthreeback(1:en
d,3);
        else
            jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,4).onset=[0.1]
        end

        if ~isempty(RespVect(1,v.jid).CorrectResp)

jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,5).onset=RespVect(1,v.jid).CorrectResp(1:en
d,3);
        else
            jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,5).onset=[0.1]
        end

        if ~isempty(RespVect(1,v.jid).AllErrors)

jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,6).onset=RespVect(1,v.jid).AllErrors(1:end,
3);
        else
            jobs{1,1}.stats{1,1}.fmri_spec.sess.cond(1,6).onset=[0.1]
        end


        BehavData(v.jid).status = 'completed spec, about to run jobman';

        spm_jobman('run',jobs);

        BehavData(v.jid).status = 'completed specification jobman';


end
```

Appendix A4: NBACK_estimate.m

```
function a = NBACK_estimate(basedir, SubjectIDs)
%STEP 3: ESTIMATE
%now estimate.  No need to import template, simple jobs file

%jobs must be cleared from previous stages
clear jobs

%run estimate loop

for i=1:size(SubjectIDs{1,1},1);

    v.jid=SubjectIDs{1,1}(i);
    v.imagedir = strcat(basedir, num2str(v.jid), '/', 'SPM.mat');
    disp(v.imagedir)

    jobs{1,1}.stats{1,1}.fmri_est.spmmat{1,1} = v.imagedir;
    jobs{1,1}.stats{1,1}.fmri_est.method.Classical = 1;
    spm_jobman('run',jobs);

end
```

Appendix A5: NBACK_Addcontrasts

```
function a = NBACK_AddContrasts(contrastfile, directory, SubjectIDs)


%SCRIPT FOR Adding Contrasts - Fergus Kane
% contrastfile = a template contrast file
% directory = base directory under which subjects directories lie
% SubjectIDs = matrix with two columns, first with subject directory (numerical),
second with subject ID


% Go to Project Directory
    cd(directory);

%Set up loop to run analysis for all subjects
%Loads a jobs file presaved from SPM batch mode and alters the SPM.mat directory

    for i=1:size(SubjectIDs{1,1},1);

         id=SubjectIDs{1,1}(i);
         load(contrastfile);
         matfile = strcat(directory, '/', num2str(id), '/', 'SPM.mat');
         jobs{1,1}.stats{1,1}.con.spmmat{1,1}= matfile;
         spm_jobman('run',jobs);

    end
```

Appendix A6: voiextractor.m

```
%  _    _  _____ _____                 _                 _
% | |  | ||  _  |_   _|               | |               | |
% | |  | || | | | | |      _____  _| |_ _ __ __ _  ___| |_ ___  _ __
% | |  | || | | | | |     / _ \ \/ / __| '__/ _` |/ __| __/ _ \| '__|
% \ \_/ /\ \_/ /_| |_  |  __/>  <| |_| | | (_| | (__| || (_) | |
%  \___/  \___/ \___/   \___/_/\_\\__|_|  \__,_|\___|\__\___/|_|
%

% This is simply a script to extract multiple VOIs from multiple subjects.

% Developed by Fergus Kane with assistance from Simone Reinders  12/2007
% If you use this script please let us know (f.kane@iop.kcl.ac.uk).
% If you have problems, let us know... but please try and work it out
% yourself first.

% NOTE WELL:  You use this script entirely at your own risk.  The are NO
% guarantees

%Does the following:
    %1. Takes a subject who has been proprocessed and has had contrasts
    %   defined.
    %2. Opens a specified contrast (effectofinterest by default).
    %3. Selects a specific voxel, runs a small volume correction to find
    %   the local maximum.
    %5. Extracts a VOI from the local maximum and saves it.
    %6. Saves a text file for each ROI, reporting successful/failed
    %   extraction

%Has the following dependencies (beyond the standard SPM5 library:

    %spm_results_ui_FK      Opens the selected contrast and calls spm_getSPM_FK
    %spm_getSPM_FK
    %spm_VOI_FK:            Runs the Small Volume Correction
    %spm_regions_FK         Generates the VOI.
    %ReadSubjectID1Column   Reads subject IDs / directories from a text file

    %In the main, these are altered versions of SPM orginals, with the GUI
    %stripped out.

%-----------------------------------------------------------------------

        %path ('/biofs/software/system/spm/spm5/toolbox/marsbar/spm5', path);
%probably not needed..
        path ('/biofs/software/system/spm/spm5/', path); %Change to SPM location.
    path ('/home/spnefek/matlab/PPI', path);
    path ('/home/spnefek/matlab/', path); %Change this to wherever the scripts are



%CALL SPM
    clear all;
    spm fmri;
        spm_defaults;
        global defaults;



%SET COMMON VARIABLES
    %setcontrast to use (must be added to each subject beforehand)
        v.contrastname='effectofinterest';
    %set base directory (where subjects directories found):
        v.basedir = '/home/spnefek/NBACK/Standard3Condition/';

    %set directory holding subjects text file
        v.subjectsdir = '/home/spnefek/matlab/NBACK/responsefiles/';
    %set subjects text file name
        v.fn = 'AllSubjectsTWINBIP.prn';
```

```
%Read Subject IDs
    Data.SubjectIDs = ReadSubjectID1Column(v.subjectsdir, v.fn);
    Data.SubjectIDs = {118}

%Specify ROIS to be extracted

Data.ROIs{1,1} = [-40;-48;46];
Data.ROIs{1,2} = [44;-46;44];
Data.ROIs{1,3} = [32;02;50];
Data.ROIs{1,4} = [06;20;50];
Data.ROIs{1,5} = [-28;00;58];
Data.ROIs{1,6} = [34;24;-06];
Data.ROIs{1,8} = [-34;22;-02];

%specify ROI names for output file


Data.ROIs{2,1} = 'VOI_A'
Data.ROIs{2,2} = 'VOI_B'
Data.ROIs{2,3} = 'VOI_C'
Data.ROIs{2,4} = 'VOI_D'
Data.ROIs{2,5} = 'VOI_E'
Data.ROIs{2,6} = 'VOI_F'
Data.ROIs{2,7} = 'VOI_G'


%Specify number of ROIS
    v.n=1


    %LOOP FOR MULTIPLE ROIS
    for j=1:v.n

    %specify the voxel to start SVC
    v.preselectedvoxel = Data.ROIs{1,j};
    v.preselectedvoxelname = Data.ROIs{2,j};


        %LOOP FOR MULTIPLE SUBJECTS
        for i=1:size(Data.SubjectIDs{1,1},1);

            v.jid=Data.SubjectIDs{1,1}(i)
            v.str = strcat(v.basedir, num2str(v.jid), '/');
            v.workingdirectory = v.str;
            Data.results{i,1}=num2str(v.jid);

        %Use try to intercept errors without crashing whole loop, stage reached
        %before crash is recorded in output file (from Data.results)

         try

            %First Open the F Contrast
                %set threshold
                v.inputthreshold=0.9;
                %call spm_results_ui, which in turn calls spm_getSPM_FK, passing on
                %the variables defined above
                [hReg,xSPM,SPM] =
spm_results_ui_FK(v.workingdirectory,v.contrastname,v.inputthreshold);
                Data.results{i,2}='spm results completed';


            %Now Run SVC to select local peak

                %specify the radius of the search sphere
                v.svcradius = 6;

                %Call SVC function
                svc = spm_VOI_FK(SPM,xSPM,hReg,v.svcradius,v.preselectedvoxel);
                %the peak voxel will be svc.dat{1,11}
                Data.results{i,2}='svc completed';

                if isempty(svc.dat)==1
                    Data.results{i,4}='No suprathreshold voxels';
                else
                    Data.results{i,4}=num2str(svc.dat{1,11}');
                end


            %Now Generate VOI from the local peak
```

```
                %sphere and radius of 8 set in function

                %Use prselected Coordinates for name

%v.name=strcat('ForDCM',num2str(v.preselectedvoxel(1)),num2str(v.preselectedvoxel(2)),
num2str(v.preselectedvoxel(3)));
                v.name = v.preselectedvoxelname
                %set radius of VOI sphere
                    v.voiradius=6;
                    v.preselectedvoxelb=svc.dat{1,11};
                    %output{i}=svc.dat{1,11};

spm_regions_FK(xSPM,SPM,hReg,v.preselectedvoxelb,v.name,v.contrastname,v.voiradius);
                    Data.results{i,2}='VOI generated';

                %copy voi file into subdirectory
                mkdir(strcat(v.workingdirectory, v.preselectedvoxelname));
                v.sourcefile = strcat(v.workingdirectory, 'VOI_',
v.preselectedvoxelname, '_1.mat');
                v.target = strcat(v.workingdirectory, v.preselectedvoxelname);
                copyfile(v.sourcefile, v.target);

                Data.results{i,3}='passed';

            catch

                Data.results{i,3}='failed';
        end

    %Write report file, check this to see if any runs failed.  Also gives
    %Coordinates of the VOI
    cd(v.basedir);
    v.resultfile =
strcat('Report',num2str(v.preselectedvoxel(1)),num2str(v.preselectedvoxel(2)),num2str(
v.preselectedvoxel(3)),'.txt');
    fid = fopen (v.resultfile, 'wt');
    v.outputdata = Data.results';
    fprintf(fid, '%s %s %s %s  \r' , v.outputdata{:} );
    fclose(fid);

    end

    end
```

Appendix A7: ReadSubjectID.m

```
function SubjectIDs = ReadSubjectID(subjectsdir, fn)

%Just a simple function to load subject directories and IDs into a matrix.
%   subjectsdir is the directory with the subjects file
%   fn is the subject file name (two columns (dir id) white space seperated)
%   !!!Note however, that for the moment, the directory must be a number!!!

    cd(subjectsdir);
    fid=fopen(fn);
    SubjectIDs=textscan(fid, '%n');
    fclose(fid);
```

Appendix A7: PPI_script_from_AM_VerFK.m

```
% call SPM to have the graphics windows in place (#mod#)
%========================================================================
%add path for spm_get in spm5
path ('/biofs/software/system/spm/spm5/toolbox/marsbar/spm5', path)
path ('/biofs/software/system/spm/spm5/', path)
path ('/home/spnefek/matlab/NBACK/', path)
path ('/home/spnefek/matlab/', path)

spm fmri
spm_defaults
global defaults
fs = filesep; % platform-specific file separator


n_sess          = 1;
n_scans         = [270];  % scans per session

% names & paths
dir_base        = '/home/spnefek/NBACK/Standard3Condition/';
dir_analysis    = 'PPI_VOI_E';
ppifilename     = 'PPI_PPIForVOI_E.mat'
%dir_functional  = {'scans'}; % base directory of functional scans (Analyze)


% subject-specific variables (#mod#)
%========================================================================

    %Read Subjects Text File
    %set directory holding subjects text file
        v.subjectsdir = '/home/spnefek/matlab/NBACK/responsefiles/';
    %set subjects text file name
        v.fn = 'AllSubjectsTWINBIPc.prn';

    %Convert Data so readable by script, script also edited to read number
    %array rather than character cell array

    ns=ReadSubjectID1Column(v.subjectsdir, v.fn);
    ns = ns(1:end,1);
    ns = ns{1:end}';
    name_subj = ns;

    %name_subj = {'10','17','18','19','20','21'}


k       = 1;

for k = [1:length(name_subj)]
    results(k,1)=name_subj(k);
    start = clock;

    v.dir = ([dir_base fs num2str(name_subj(k)) fs dir_analysis]);

    cd(v.dir);
      %  cd ([dir_base fs name_subj{k});


    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    SPM.nscan          = [270];  % con 2 sessioni con un numero di scans diverso, [120
130]

    % Here one has to choose the basis function used:
    %-----------------------------------------------------------------------
    % OPTIONS:'hrf'
    %         'hrf (with time derivative)'
    %         'hrf (with time and dispersion derivatives)'
    %         'Fourier set'
    %         'Fourier set (Hanning)'
    %         'Gamma functions'
    %         'Finite Impulse Response'
    %-----------------------------------------------------------------------
```

```
    %In this example, one choses the hrf (with time derivative):
    SPM.xBF.name       = 'none';

    %The following default values don't need to be changed apart from SPM.xBF.UNITS:
    %here one has to write 'secs' instead of 'scans' if the stimulus onsets in
    %SPM.Sess(1).U(1).ons are expressed in seconds instead of scans:

    SPM.xBF.length     = 32;               % length in seconds
    SPM.xBF.order      = 1;                % order of basis set
    SPM.xBF.T          = 16;               % number of time bins per scan
    SPM.xBF.T0         = 8;                % first time bin (see slice timing)
    SPM.xBF.UNITS      = 'secs';           % OPTIONS: 'scans'|'secs' for onsets
    SPM.xBF.Volterra   = 1;                % OPTIONS: 1|2 = order of convolution
    SPM.xY.RT          = 2;                % seconds




    SPM.Sess(1).U   = [];



    %Here one can add user specified covariates. In the example, a 'Regressor 1' made
by a vector [1:360] is used.
    %In matlab, the expression [1:360] indicates a vector [1 2 3 4 5... 358 359 360]
which includes 360 numbers.
    %-----------------------------------------------------------------------

    load(ppifilename)

    SPM.Sess(1).C.C    = PPI.ppi;          % [n x c double] covariates
    SPM.Sess(1).C.name = {'PPI regressor'};   % [1 x c cell]   names

    % global normalization: OPTINS:'Scaling'|'None'
    %-----------------------------------------------------------------------
    SPM.xGX.iGXcalc    = 'None';

    % low frequency confound: high-pass cutoff (secs) [Inf = no filtering]
    %-----------------------------------------------------------------------
    SPM.xX.K.HParam    = Inf;

    % intrinsic autocorrelations: OPTIONS: 'none'|'AR(1) + w'
    %-----------------------------------------------------------------------
    SPM.xVi.form       = 'none';

    % specify data by matrix of filenames (#mod#)
            %-------------------------------------------
                SPM.xY.P = [];


                for sess    = 1:n_sess,

                    dir_scans{sess} = [dir_base fs num2str(name_subj(k)) ];
                    cd(dir_scans{sess});

                    % load scan file names into cell array (#mod#)
                    Filter             = 'sw*.img';
                    nscan              = SPM.nscan(sess);
                    temp{sess}  = spm_get('files', dir_scans{sess}, Filter); %NB:
change the prefix to 'ssw' if batch script was used for preprocessing
                    %trying spm_select instead
                    %temp{sess}  = spm_select('files', dir_scans{sess}, Filter); %NB:
change the prefix to 'ssw' if batch script was used for preprocessing
                    temp{sess}          = temp{sess}(1:nscan,:);
                    SPM.xY.P           = char(temp);
                end

                % concatenate scans across all sessions
                % (spm_fmri_spm_ui expects that all file names are in SPM.xY.P)



     % cd ([dir_base fs name_subj{k} fs dir_analysis]);
     cd ([dir_base fs num2str(name_subj(k)) fs dir_analysis]);

    % Configure design matrix
    %=======================================================================
    SPM = spm_fmri_spm_ui(SPM);
```

```
    % Estimate parameters
    %===========================================================================
    SPM = spm_spm(SPM);



    %k= k + 1;
    % set err to zero and clear major variables

    clear SPM;



        load SPM
%
%       % Add extra contrasts
%       %=========================================================================
        w = length(SPM.xCon);
%
%       % T-contrasts
% %     %----------------------------------------------------------------------
%
%
%
%     %%%%%%%%%%%%%%%%%%%%%RHYME%%%%%%%%%%%%%%%%%%%%%%%%%%
%
        c                   = [1 0];
        cname               = 'pos corr';
        SPM.xCon           = spm_FcUtil('Set',cname,'T','c',c',SPM.xX.xKXs);
%
%       c                   = [0 0 0 0 0 0 1 0 0 -1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
%       cname               = 'simple effect AL > TN CORRECT';
%       SPM.xCon(end + 1)   = spm_FcUtil('Set',cname,'T','c',c',SPM.xX.xKXs);
%
%
%     %%%%%%%%%%%%%%%%%%%%%%READ%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%       c                   = [0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 -1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
%       cname               = 'simple effect reading words > FF ';
%       SPM.xCon(end + 1)   = spm_FcUtil('Set',cname,'T','c',c',SPM.xX.xKXs);
%
%
%       c                   = [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 -1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
%       cname               = 'simple effect reading non words > FF';
%       SPM.xCon(end + 1)   = spm_FcUtil('Set',cname,'T','c',c',SPM.xX.xKXs);
%
%
%       % and evaluate
%       %---------------------------------------------------------
        spm_contrasts(SPM, w+1:length(SPM.xCon));
%
%

     %k= k + 1;
%     % set err to zero and clear major variables
%
     clear SPM;
%
 finish=clock;
results(k,2)=etime(finish,start);
end
```

8.2. Appendix B: List of Abbreviations

Abbreviation	Definition
ACC	Anterior Cingulate Cortex
ADC	Apparent Diffusion Coefficient
AF	Arcuate Fasciculus
AgCC	Agenesis of the Corpus Callosum
ALIC	Anterior Lateral Interior Capsule
ASRM	Altman Self Rating Scale
BD	Bipolar Disorder
BD-I	Bipolar Disorder I
BD-II	Bipolar Disorder II
BDI	Beck Depression Index
BOLD	Blood Oxygen Level Dependent
CANTAB	Cambridge Neuropsychological Test Automated Battery
CC	Corpus Callosum
CVLT	California Verbal Learning Test
DB	Discordant Bipolar
DCM	Dynamic Causal Modeling
DLPFC	Dorsolateral Prefrontal Cortex
DNB	Discordant-Non-Bipolar
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
DTI	Diffusion Tensor Imaging
DZ	Dyzygotic (Fraternal)
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
fMRI	Functional Magnetic Resonance Imaging
GFAP	Glial Fibrillary Acidic Protein
HamD	Hamilton Depression Scale
IFG	Inferior Frontal Gyrus
IHT	Interhemispheric Transfer
ILF	Inferior Longitudinal Fasciculus
MBTS	Maudsley Bipolar Twin Study
MD	Mean Diffusivity
medPFC	Medial Pre Frontal Cortex
MFG	Middle Frontal Cortex
MRI	Magnetic Resonance Imaging
MZ	Monozygotic (Identical)
OCD	Obsessive Compulsive Disorder
OFC	Orbital Frontal Cortex
PCC	Posterior Cingulate Cortex
PET	Positron Emission Tomography
PFC	Pre Frontal Cortex
PPI	Psychophysical Interaction
PSC	Parental Social Class

rCBF	Rate of Cerebral Blood Flow
ROI	Region of Interest
SCAN	Schedules of Clinical Assessment in Neuropsychiatry
SFG	Superior Frontal Gyrus
SG	Subgenual Cingulate Gyrus
SLF	Superior Longitudinal Fasciculus
SPET	Single Photon Emission Tomography
SPM	Statistical Parametric Mapping
TBSS	Tract Based Skeletal Statistics
UF	Uncinate Fasciculus
UR	Unaffected Relatives
VBM	Voxel Based Morphometry
VLPFC	Ventrolateral Prefrontal Cortex
VOI	Volume of Interest
WASI	Wechsler Abbreviated Scale of Intelligence
WCST	Wisconsin Card Sorting Test
WHO	World Health Organisation
YMRS	Young Mania Rating Scale